The phylogenetic and phenotypic analysis of *Salmonella enterica* serovar Weltevreden

Eugenie Carine MAKENDI NJOCK

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of London

September 2015

Department of **Pathogen Molecular Biology**

Faculty of **Infectious and Tropical Diseases**

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by **NOVARTIS INSTITUTE FOR TROPICAL DISEASES**

Research group affiliation(s): Dougan Faculty **Microbial Pathogenesis: Wellcome Trust Sanger Institute**

Abstract

Diarrhoeal diseases remain a global health threat and are responsible for high levels of morbidity and mortality worldwide, with an estimated 1.7 billion cases every annum. Additionally, according to the World Health Organisation, diarrhoeal diseases are the second leading cause of death in children under 5 years old. *Salmonella* are one of the most common diarrhoeal pathogens [1] (WHO Accessed 20 February 2015) with serovars Enteritidis, Typhimurium and Typhi playing a major role in outbreaks worldwide. However, *Salmonella enterica* serovar Weltevreden (*S.* Weltevreden) has recently attracted a great deal of interest due to increasing reports of its isolation by reference laboratories around the world, with a particular high incidence in South East Asia. However, relatively little is known about the genotypic or phenotypic properties of this understudied serovar.

In this study, phylogenetics and comparative genomics based on whole genome sequences were used to define the genetic diversity within a sizeable collection of *S.* Weltevreden isolates collected from across the globe, with a focus in South East Asia. This phylogenetic analysis confirmed that the *S.* Weltevreden isolates belong to a monophyletic clade formed of several sub-clades presenting distinct geographical clustering and characteristics. Phenotypic characterisation was performed on selected isolates, with an aim to dissect aspects of host-pathogen interaction during infection, providing a foundation to compare *S.* Weltevreden with other serovars such as *S.* Typhimurium. Interestingly, an overall attenuated pathology was observed both *in-vitro* (hep 2 cell line) and *in-vivo* (murine and zebrafish embryos) for *S.* Weltevreden compared to the *S.* Typhimurium reference strain.

This is the first report of the phylogenetic analyses of *S.* Weltevreden and of a systematic *in-vitro* and *in-vivo* characterisation of the sub-species.

Declaration

The work reported in this thesis was conducted from May 2012 to September 2015 at the Wellcome Trust Sanger Institute (WTSI) under the supervision of **Professor Gordon Dougan** (WTSI, Cambridge UK) and **Professor Brendan Wren** (London School of Hygiene and Tropical Medicine (LSHTM), London UK).

This thesis is the result of my own work and contains nothing that is the outcome of third parties' work or collaboration except where specifically acknowledged. All reported results and quotations from the published or unpublished work of other people has been recognised and appropriately acknowledged.

All experimental procedures and data analysis were performed by myself except for the mice infection challenges that were performed by **Dr. Simon Clare** (WTSI), the zebrafish injections were performed by **Dr. Antonio Pagan** and **Dr. Steven Levitte** (MRC Laboratory of Molecular Biology (LMB), University of Cambridge, Cambridge UK). The microscopy pictures presented in this thesis were taken by **David Goulding**. The DNA samples were sequenced, assembled and annotated with the **WTSI Sequencing** teams and the **Pathogen Informatics** team. All scripts used for the phylogenetic studies were written and developed mainly by the **Dr. Simon Harris** (WTSI) and members of the **Pathogen Informatics** team (WTSI). Assistance, guidance and advices from Senior Scientists were provided throughout the duration of the project.

This thesis does not exceed 100.000 words in length, including tables, figures and footnotes in accordance to the school regulations.

Signature: ……..

Date: ……

Acknowledgements

I would like to express my sincere gratitude to **Professor Gordon Dougan** (WTSI, Cambridge UK) for the opportunity to work in his team, his patient guidance, his immense knowledge, and for the continuous support and motivation throughout this journey. My sincere thanks also go to **Professor Brendan Wren** (LSHTM, London UK) for the dedicated supervision and mentoring, for his insightful comments and encouragements.

Beside my supervisors, I am beholden to a number of key people for making this work possible, particularly **Dr. Stephen Baker** (Enterics Infection, OUCRU, Ho Chi Minh, Vietnam) for providing the core concept of the project.

I would like to offer my special thanks to **Dr. Francois Xavier Weill** and **Dr. Simon le Hello** (Institut Pasteur, Paris France) as well as **Dr. Tu Le Thi Phuong** and **Dr. Stephen Baker** (Enterics Infection, OUCRU, Ho Chi Minh, Vietnam) whose collaborative effort contributed to the worldwide sample collection used in this study.

I would like to express my very great appreciation to **Dr. Andrew Page** (WTSI, Cambridge UK) for the invaluable guidance, help and assistance during the phylogenetic analysis and for reviewing and proof reading sections of this thesis. I am thankful for the help provided by the **Pathogen Informatics** and **Sequencing teams** of the WTSI, Cambridge UK for sequencing and assembling the samples.

The patient guidance of **Theresa Feltwell** (WTSI, Cambridge UK) has been a great help in microarray experiments.

I am particularly grateful for the assistance and advices provided **by Dr. Derek Pickard** (WTSI, Cambridge UK) on the microbiological characterisation of the isolates used in this study.

I wish to acknowledge the help provided by **Dr. Christine Hale** (WTSI, Cambridge UK) in designing the cellular assays and analysing the outcome.

Assistance provided by **David Goulding** (WTSI, Cambridge UK) with imaging was greatly appreciated.

My special thanks goes to **Dr. Simon Clare** (WTSI, Cambridge UK), for performing the infection challenges in mice and providing support in analysing the data.

I am indebted to **Dr. Antonio Pagan** and **Dr. Steven Levitte** (LMB, University of Cambridge, Cambridge UK) for their major contribution in zebrafish infections.

A special thanks to **Sophie Palmer and Kate Auger** (WTSI, Cambridge UK) and **Helen White** (LSHTM, London UK) for patiently dealing with all administrative matters.

My special thanks are extended to all members of **Professor Gordon Dougan's** Faculty Team (WTSI, Cambridge UK) particularly, **Dr. Elizabeth Klemm, Dr. Vanessa Wong, Dr. Robert Kingsley, Dr. Amy Yeung, Leanne Keane** and **Sally Kay**, for the stimulating discussions, help and support with experiments, troubleshooting and logistics.

Last but not the least, I am most grateful to **Novartis Institute for Tropical Diseases** (NITD) for providing the funding for this research and supplying the tuition fees and living expenses throughout the duration of the project. My sincere thanks to **Professor Paul Herrling** (Novartis, Basel Switzerland) for his encouragements and guidance and **Mrs. Patsy Tan** (NITD, Singapore) for patiently dealing with all the administrative procedure from overseas.

Table of Contents

Abstract ……..2

Declaration ……3

Acknowledgments ……4

Table of content ……6

List of figures ……9

List of tables ……12

Glossary ……13

1 Introduction 15

1.1 Diarrhoeal diseases 16

1.2 Salmonella enterica 19

1.2.1 Typhoidal *Salmonella* 22

1.2.2 Nontyphoidal *Salmonella* (NTS) 24

1.3 Molecular approaches to identify and discriminate between S. enterica 25

1.4 Evolution of the genus Salmonella 29

1.5 Major virulence-associated genes 34

1.5.1 Virulence-associated plasmid 38

1.5.2 Examples of other virulence-associated genes 38

1.6 Signatures of adaptation in S. enterica 39

1.6.1 Genome degradation/decay 42

1.7 Methods for phenotyping 43

1.7.1 *In-vitro* models 43

1.7.2 *In-vivo* model: The mouse 44

1.7.2.1 The streptomycin-pretreated mouse model for colitis 46

1.7.3 *In-vivo* model: The zebrafish 46

1.7.4 The mouse genetic screening 50

1.7.5 The zebrafish genetic screening 50

1.8 The use of whole genome sequencing for studying bacterial genomes and phylogeny 51

1.9 The focus of this thesis, S. Weltevreden 54

1.10 Aims and objectives of thesis 54

2 Materials and methods 56

2.1 Phylogenetic analysis of S. Weltevreden 56

2.1.1 Illumina sequencing 56
2.1.2 Sequence assembly from Illumina reads 57
2.1.3 Pacific Biosciences (PacBio) assembly 58
2.1.4 Annotation 60
2.1.5 MLST from *de novo* assemblies 60
2.1.6 Checking for *S*. Weltevrden in sequencing reads 60
2.1.7 Detecting regions likely to be erroneous with short read sequencing 61
2.1.8 Recombination mapping 61
2.1.9 General mapping 62
2.1.10 SNP calling 62
2.1.11 Clusters and defining SNPs 63
2.1.12 Predicting antibiotic resistance 63
2.1.13 Pan genome analysis 64
2.2 Phenotypic characterisation of S. Weltevreden 65
2.2.1 Bacterial strains culture conditions 65
2.2.2 Serological identification 65
2.2.3 Bacterial growth assessment 66
2.2.4 Microarray assay (Biolog) 66
2.2.5 Gentamicin killing assays using Hep2 cells 67
2.2.6 Confocal microscopy 67
2.2.7 Scanning Electron Microscopy 68
2.2.8 Transmission Electron Microscopy 68
2.2.9 Murine intravenous challenge 69
2.2.10 Colitis infection challenges 69
2.2.11 Histology 69
2.2.12 The zebrafish challenge model 70
3 Phylogenetic diversity within *S.* Weltevreden 71
3.1 Introduction 71
3.2 Results 76
3.2.1 The *S*. Weltevreden collection 76
3.2.2 Generation of a *S*. Weltevreden reference genome 78
3.2.3 Phylogenetic analysis 80
3.2.3.1 Confirmation of sequence type 80
3.2.3.2 *S*. Weltevreden in the context of other *S. enterica* 80
3.2.4 Genetic diversity of *S*. Weltevreden 86
3.2.4.1 Chromosome analysis 87
3.2.4.2 Plasmids 93
3.2.4.3 Resistome 95
3.2.4.3.1 Additional plasmids present in the samples flagged as antimicrobial resistant 95
3.2.4.4 Accessory genome analysis 97
3.2.4.5 Phylogeny of the distribution of phage-like elements 102
3.3 Discussion 107
4 Phenotypic characterisation of *S*. Weltevreden 109
4.1 Introduction 109
4.2 Results 110
4.2.1 Microbial characterisation and confirmation of serotype 110
4.2.2 Bacterial growth *in-vitro* 112

4.2.3 Metabolic profiling using the Biolog Phenotype Microarray system 112
4.2.4 Invasion into Hep 2 cells 114
4.2.4.1 Electron microscopy 118
4.2.5 *S.* Weltevreden in the murine model 122
4.2.5.1 Systemic challenge 122
4.2.5.2 Evaluation of *S.* Weltevreden in the streptomycin pre-treated colitis model 123
4.3 Discussion **127**

5 *S.* Weltevreden in the zebrafish infection model 130

5.1 Introduction **130**

5.2 Results **132**

5.2.1 Bacterial growth *in-vitro* at 28°C 132
5.2.2 Infection challenge 133
5.2.3 *Salmonella* viability in the zebrafish embryo model 134
5.2.4 Survival of *S.* Weltevreden within macrophage-deficient embryos 136
5.3 Discussion **138**

6 Summary and future directions 140

7 References 148

Appendix 1 172

Appendix 2 183

Appendix 3 188

Appendix 4 197

Appendix 5 206

Appendix 6 212

List of figures

Figure 1.1: Incidence of pathogen-specific moderate-to-severe diarrhoea per 100 child-years 17
Figure 1.2: Phylogeny of selected members of the enterobacteriaceae based on sequence comparison of core genes within the respective genomes 20
Figure 1.3: General overview of the current classification of *S. enterica.* Taken from [59] 26
Figure 1.4: Maximum likelihood phylogenetic tree of *Salmonella* based on concatenated MLST loci. Taken from [71] 30
Figure 1.5: Diagram summarising selected aspects of the evolutionary history of *S. bongori* and *S.* Typhi, a comparative member of *S. enterica.* Taken from [71] 33
Figure 1.6: Overview of the development of the zebrafish immune system. Taken from [165]. 48
Figure 2.1: a. graphic representation of S. Weltevreden genome assembly before manual fixing. b. Graphic representation of S. Weltevreden assembly before and after manual fixing. 59
Figure 3.1: Phylogenetic tree generated using the core genes (~2650 coding sequences) of various *Salmonella.* Taken from [238] 73
Figure 3.2: Mauve progressive alignment of the draft genomes of *S.* Weltevreden 2007-60-3289-1, S. Dublin CT_02021853 and *S.* Weltevreden SL484 75
Figure 3.3: Geographical distribution of *S.* Weltevreden isolates included in the study 77
Figure 3.4: DNA plotter diagram of *S.* Weltevreden 10259 genome 79
Figure 3.5: Maximum likelihood tree comparing *S.* Weltevreden with selected *S. enterica* serovars 83
Figure 3.6: Simplified phylogeny of selected *S. enterica* serovars deposited on the NCTC database compared to *S.* Weltevreden. 86
Figure 3.7: Gubbins-based analysis representing the recombination events across the tree 89
Figure 3.8: Population Structure of *S.* Weltevreden isolates with key metadata information 93

Figure 3.9: Maximum likelihood plasmid tree build on 48 SNPs..... 94
Figure 3.10: Breakdown of the frequency of gene in isolates and in the overall collection of *S.* Weltevreden 99
Figure 3.11: Plot showing variance in the number of unique genes found in one isolate only and the number of new genes as genomes are added to the pan genome. 100
Figure 3.12: Variance in the total number of predicted CDSs (genes) in the pan genome and the of conserved CDSs (99% of isolates) in the core genome as samples are added. 101
Figure 3.13: Predicted phage distribution across the genome of the *S.* Weltevreden 10259..... 103
Figure 3.14: Comparative analysis of the genome sequence of *S.* Weltevreden (1) 10259 (2) 98_11262 (3) 99_3134 and (4) C2346 106
Figure 4.1: Standarisation curve reporting the number of colony forming units (CFUs) per OD600 in LB medium for each bacterial isolate 112
Figure 4.2:Carbon source utilisation microarray 113
Figure 4.3: Confocal microscopy of *S.* Typhimurium SL1344 and *S.* Weltevreden in Hep 2 cells, 2 and 6 hours post exposure..... 116
Figure 4.4: Number of viable *Salmonella* recovered in gentamicin killing assay ... 118
Figure 4.5: SEM of Hep 2 cells infected with *S.* Typhimurium SL1344 or *S.* Weltevreden C2346..... 120
Figure 4.6: Typical TEM ultrastructures visualised within Hep 2 cells infected with *S.* Typhimurium (panel A) or *S.* Weltevreden (panel B)..... 121
Figure 4.7: Percentage of survival of C57bl/6 mice challenged with *S.* Typhimurium SL1344, ... 123
Figure 4.8: Histopathological analysis of caecum sections of mice infected with *S.* Typhimurium SL1344, *S.* Weltevreden C2346 or mock-injected with PBS 4 days post infection..... 125
Figure 4.9: Analysis of the impact of colonisation of the ceacum by *S.* Typhimurium SL1344, ... 126
Figure 4.10: Analysis of the levels of liver colonisation of *S.* Typhimurium SL1344, *S.* Weltevreden C2346 or *S.* Weltevreden 10259 in C57bl/6 mice 4 days post infection..... 127

Figure 5.1: Transgenic zebrafish embryos infected with *S.* Typhimurium SL1027 48 hpi. Taken from [164] 131
Figure 5.2: Growth curve showing the duplication time of *S.* Typhimurium SL1344 , *S.* Weltevreden C2346 and *S.* Weltevreden 10259 in LB medium at 28°C. 132
Figure 5.3: Percentage of survival of zebrafish embryos microinjected with *S.* Typhimurium SL1344 (ST), *S.* Weltevreden C2346 (SW) or mock injected 133
Figure 5.4: Analysis of Salmonella survival within zebrafish embryos microinjected with 250 - 300CFUs of *S.* Typhimurium SL1344 or *S.* Weltevreden C2346 134
Figure 5.5: Analysis of the replication level within zebrafish embryos microinjected with 250 - 300 CFUs of *S.* Typhimurium SL 1344 (ST SL1344) and S. Weltevreden C2346 (SW C2346) 20 hours post infection - P value < 0.0001. 135
Figure 5.6: Kinetics of *S.* Weltevreden C2346 survival within zebrafish embryos over the course of the infection following challenge with 250 - 300 CFUs. 136
Figure 5.7:Percentage of survival of wild-type irf8+/+, heterozygous mutants irf8st95/+ and homozygous mutants irf8 st95/st95 challenged with *S.* Typhimurium SL 1344 (ST), *S.* Weltevreden C2346 (SW) or mock-injected 70 hours post infection 137
Figure 6.1: Circular map of *S.* Weltevreden 10259 plasmid 143

List of Tables

Table 3.1: MLST data for *S.* Weltevreden 10259 reporting alleles numbers. 80
Table 3.2: List of *S. enterica* serovars and Isolates used for comparison and their accession numbers where available 85
Table 3.3: Genes known to confer antimicrobial resistance found in *S.* Weltevreden and the isolates in which they were detected. 95
Table 3.4: Isolates likely to harbour antimicrobial resistance-associated plasmids... 97
Table 4.1: Sera agglutination results summary 111

Glossary

AMR	Antimicrobial Resistance
ATP	Adenosine Triphosphate
BAPS	Bayesian Analysis of Population Structure
bp	Base pairs
C.	Campylobacter
CDS	Codind DNA Sequence
CFUs	Colony Formin Units
CRISP	Clustered Regularly Interspaced Short Palindromic Repeats
ddNTP	Dideoxyribonucleotides
DNA	Deoxyribonucleic acid
DSS	Dextran Sodium Sulphate
ENU	Ethylnitrosourea
ETEC	Enterotoxigenic Escherichia coli
GEMS	Global Enterics Multi-centre Study
GFP	Green Fluorescent Protein
Hpf	Hours post fertilisation
Hpi	Hours post Infection
LB	Lysogeny Broth
LPS	Lypopolysaccharide
LSHTM	London School of Hygiene and Tropical Medicine
MLEE	Multi-Locus Enzyme Electrophoresis
MLST	Multi Locus Sequence type
MOI	multiplicity of infection
MRC	Medical Research Council
NCTC	National Collection of Type Cultures
NRPSs	non-ribosomal peptide synthetases
NTS	Non Typhoidal Salmonella(Salmonellosis)
OUCRU	Oxford University Clinical Research Unit

PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
PK	polyketide synthases
RFLP	Restriction Fragment Length polymorphism
S.	*Salmonella*
SCV	Salmonella Containing Vacuole
SEM	Scanning Electron Microscopy
SMRT	Single Rolecule Real Time (sequencing)
SNPs	Single nucleotide polymorphisms
SPI	Salmonella pathogenicity Island
ST	Sequence Type
TALENs	Transcription activator-like effector nucleases
TEM	Transmission Electron Microscope
TILLING	Targeting Induced Local Lesions in Genomes
WHO	World Health Organisation
WTSI	Wellcome Trust Sanger Institute
ZFN	zinc-finger nuclease
ZMP	Zebrafish Mutation Project

1 Introduction

Infectious diseases are still one of the most common causes of morbidity and mortality in both humans and domestic animals. In regions of the world with poor economic development, many of the classical infections such as cholera and typhoid still persist and new diseases such as ebola are emerging [2-4]. In the more economically developed parts of the world, infections associated with zoonosis or health-care systems are an ever-present threat and given the increasing age demographic these threats are unlikely to recede [5, 6]. Further, the global emergence of multiple drug resistant microbes is challenging our ability to treat infections in a reliable manner and new antimicrobials are urgently required [7]. Thus, the ability to identify and understand pathogens and determine how they are spread will be central to our future success in combatting our old adversaries.

The challenges presented by infections are considerable. It is clear that microbes have the ability to evolve rapidly to fill new ecological niches and to resist our attempts to kill them. Fortunately, these challenges coincide with an era where we are learning more about the epidemiology of disease and the molecular basis of infection. Modern molecular and immunological techniques, linked to the genomic sciences, are providing a rich source of tools for studying pathogen genetics, immunity and evolution [8-10]. Over the past decades a reductionist approach has yielded vital information on how microbial genes contribute to virulence and how their translated products contribute to pathogenesis [11, 12]. More recently, mammalian genetics is revealing the contribution of host genes to infection susceptibility [13-15]. The introduction of new generation sequencing technologies has revolutionised our ability to sequence the genomes of microbial populations, facilitating the analysis of whole genomes as a source of information on phylogeny, pathogen evolution and disease transmission [16-18]. As more human genomic data becomes available, similar progress can be anticipated in terms of identifying infection-susceptible individuals. It is hoped that many of these basic scientific advancements will begin to have impact in the form of improved treatments in the near future.

1.1 Diarrhoeal diseases

Infections of the intestine associated with diarrhoea are still an important component of the global infectious disease burden with an estimated 1.7 billion cases of diarrhoeal disease every year [19]. In the developing world these impact disproportionately on new-borns and infants, particularly those that are poorly nourished or lack access to an effective public health infrastructure. According to the World Health Organisation (WHO), diarrhoeal diseases are the second leading cause of death in children under 5 years old and are accountable for ~760,000 deaths annually out of 2.2 million deaths overall [20].

A recent global study, known as the Global Enterics Multi-Center Study (GEMS), reported extensively on links between the incidence of severe diarrhoea and death [21]. GEMS was funded by the Bill and Melinda Gates Foundation (Gates) and managed from the University of Maryland, covering sites in sub-Saharan Africa (Kenya, Mali, Mozambique, and The Gambia), and South Asia (Bangladesh, India, and Pakistan). GEMS ran over a 3-year period involving children aged from 0 – 59 months residing in high endemic areas. The GEMS study found that the risk of child mortality was 8.5 fold higher in infants with moderate to severe diarrhoea than in children without diarrhoea. GEMS also highlighted that rotaviruses, *Cryptosporidium*, Enterotoxigenic *Escherichia coli* producing heat- stable toxin (ST-ETEC; with or without co-expression of heat-labile enterotoxin (LT)), and *Shigella* were important cases of severe to moderate diarrhoea in many of these sites. Other pathogens were frequently isolated from diarrhoea cases either at specific sites, within regions or were consistently present, but at lower incidence levels (*Aeromonas*, *Vibrio cholerae* O1, *Campylobacter, Salmonella*). Figure 1.1 shows the incidence of pathogen-specific diarrhoea per 100 children per year, in different age stratum from the GEMS data.

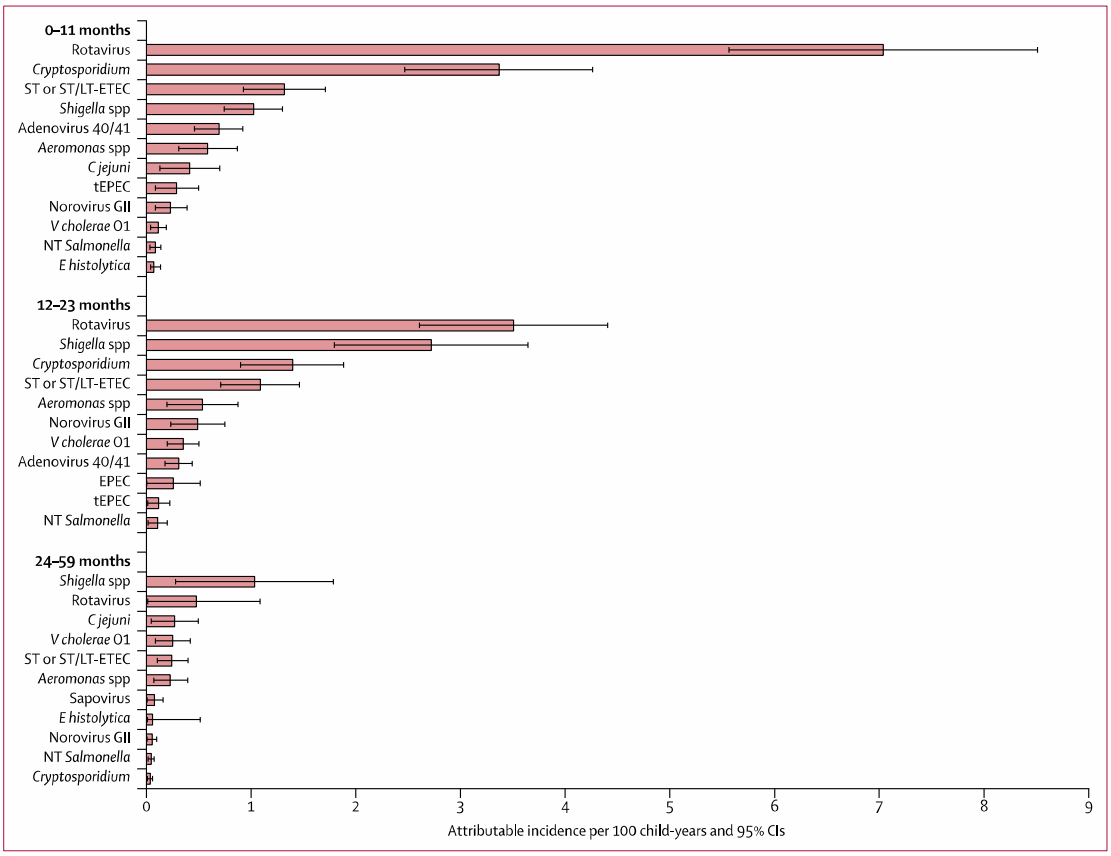


Figure 1.1: Incidence of pathogen-specific moderate-to-severe diarrhoea per 100 child-years

Incidence of pathogen-specific moderate-to-severe diarrhoea per 100 child-years by age stratum in endemic areas (all sites combined). The bars show the incidence rates and the error bars show the 95% CIs [21]. Data taken from the GEMS study.

GEMS and other related studies suggest that interventions specifically targeting the pathogens causing diarrhoea, potentially by identifying transmission pathways or reservoirs, could be a route towards significantly reducing the burden of diarrhoeal diseases. Understanding the microbial complexity of the diarrhoeal diseases could also facilitate the implementation of new prevention and treatment strategies that are urgently needed if we are to reduce infections and improve disease outcomes [21].

Salmonella were consistently isolated at the different GEMS sites, although they were clearly not the dominant cause of severe diarrhoea at any particular site. However, in addition to any role in causing diarrhoea in children, *Salmonella* can play a key role in bloodstream infection in Africa and Asia, adding to their overall disease burden. Bloodstream infection in Africa is a major healthcare threat associated with high mortality. A review on community-acquired bloodstream

infections in Africa, analysing 22 case-studies of infection, revealed that overall 13.5% of adults and 8.2% of children had experienced some form of bloodstream infection [22]. Where recorded, patients with systemic infections had an overall case fatality of 18·1%. Although malaria accounted for a significant proportion of systemic infections, *Salmonella* was responsible of ~30% of non-malaria invasive diseases (~60% of these being nontyphoidal *Salmonella* (NTS)). *Streptococcus pneumoniae* was also a significant cause of invasive bacterial disease in the same regions and was the most commonly isolated bacterial species in children. HIV co-infection was associated with an increased risk of invasive bacteraemia in general; particularly with *Salmonella enterica* and *Mycobacterium tuberculosis* [22]. A number of different serovars of *Salmonella* were associated with invasive disease (see below).

The burden of infectious diarrhoea in developed countries is generally distinct from that in the developing world. In economically developed regions, many diarrhoeal diseases are acquired through contamination of the food chain, involving zoonotic sources such as meat, dairy products or indirectly contaminated produce such as lettuces exposed to sewage [23, 24]. At a global level, *Salmonella* has been reported as one of the most common foodborne pathogens by the WHO along with *Campylobacter, E. coli, Shigella* and *Trichinella* [1]. Diseases associated with *Salmonella* are frequently caused by clades/strains that are spreading internationally (see below). *E. coli* infections, particularly those involving Enteropathogenic or Enterohemorrhagic *E. coli* (EPEC/EHEC) are also relatively common, as are *Shigella* infections in children associated with *Shigella sonnei* [25, 26]. The spread of *Shigella* is likely to involve human to human transmission rather than a zoonotic source as *Shigella* is largely a human restricted pathogen. Other common causes of diarrhoeal disease in developed regions are viruses such as norovirus and rotaviruses that also usually spread directly between humans. Perhaps the most common cause of bacterial diarrhoea in many developed countries are specific *Campylobacter* species, especially *C. jejuni* [27]. These are frequently found associated with the handling and improper cooking of poultry. In addition, travellers to endemic regions of the world can bring back cases of disease (known as traveller's diarrhoea).

Another emerging form of diarrhoeal diseases is the so-called antibiotic-associated

diarrhoeas that can follow treatment with antibiotics. Such infections are particularly common in the elderly and the immunocompromised in health-care associated settings such as hospitals [28]. A common cause of antibiotic-associated diarrhoea is due to *Clostridium difficile* which has emerged to prominence over the past few decades [29]. *C. difficile* is a broad species from which highly infectious sub-clades emerge periodically to cause epidemics [30-32]. Many of these clades are now endemic within the hospital systems of countries. *C. difficile* can overgrow in the intestine following antibiotic treatment, releasing potent enterotoxins that cause significant local and systemic pathology [33]. Individuals treated with antibiotics frequently relapse with *C. difficile* infection and thus the clinical management of the disease is challenging.

1.2 *Salmonella enterica*

Salmonella are a common cause of infections in many parts of the world in both humans and animal species. *Salmonella* infection is associated with different disease syndromes ranging from acute gastroenteritis/diarrhoea to generally more chronic systemic diseases such as typhoid. *Salmonella* are rod shaped non-spore forming and Gram-negative bacteria. They are predominantly motile due to the expression of flagellae and individual cells can vary somewhat in size. They are also chemo-organotrophs and facultative anaerobes.

The genus *Salmonella* was named after Daniel Elmer Salmon, an American veterinary pathologist who discovered the pathogen while searching for the cause of common hog cholera. *Salmonella* are *Gammaproteobacteria* and belong in the family *Enterobacteriaceae*. It is generally acknowledged that there are two species of *Salmonella*; *Salmonella bongori* and *S. enterica*. *S. bongori* is commonly found in cold-blooded animals, including reptiles and snakes and likely evolved within such animals. *S. bongori* can occasionally cause diarrhoeal disease in humans, although this is relatively rare [34, 35]. *S. enterica* is an old, broad and complex species that is likely millions of years old [36]. It harbours a number of distinct sub-species but the most common causes of disease in humans and veterinary animals fall into *S. enterica* sub-species I (Figure 1). The six main sub-species are *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV) and *indica* (VI).

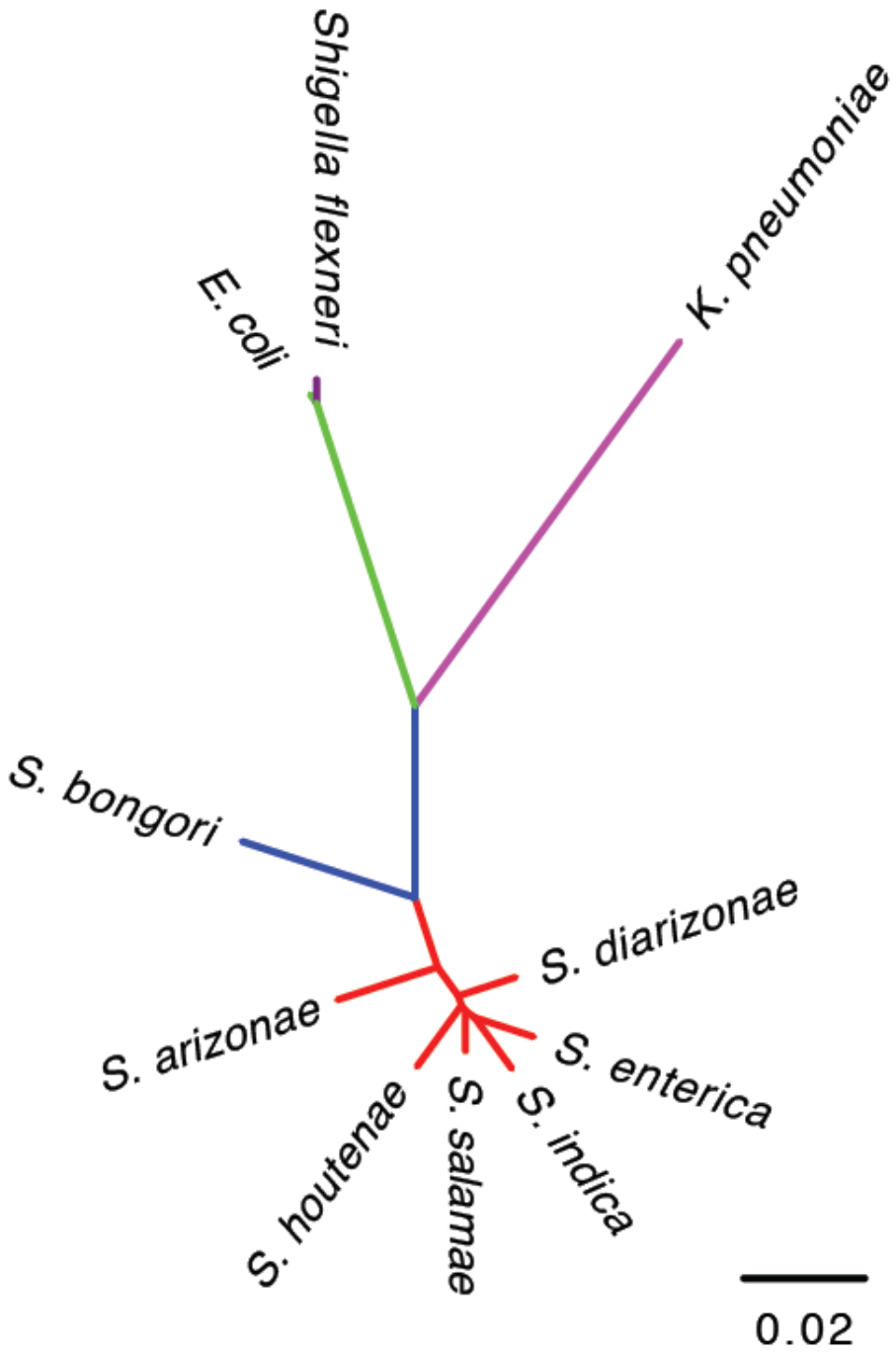


Figure 1.2: Phylogeny of selected members of the enterobacteriaceae based on sequence comparison of core genes within the respective genomes

Connecting lines coloured according to species/subspecies: six subspecies of *S. enterica* in red, *S. bongori* in blue, *E. coli* in green, *S. flexneri* in purple, and *Klebsiella pneumoniae* in pink. Branch lengths are indicative of the estimated substitution rate per variable site. Taken from. [37]

Salmonella nomenclature is complex and has changed over time. Initially, isolates were designated or named according to clinical considerations or after the host species they were frequently isolated from, for example, *S.* Typhimurium (mouse typhoid fever), *Salmonella* Cholerasuis (hog cholera) etc. Later on, it was recognised that host specificity was not a property of all Salmonella; therefore new strains or serovar (short for serological variants) were named according to the location where they were isolated from; for instance *Salmonella* Newport, *Salmonella* Montevideo.

The above nomenclature was developed without any specific consideration of the phylogeny of the isolates. Many thought that individual Salmonella serovars were actually separate species but DNA comparison studies proved that this was not the case. Since *S. enterica* can refer to either a species or subspecies a rethink on the nomenclature was required. Committees of the WHO and other esteemed bodies, alongside the recognition that *S. enterica* was a single species, developed a new nomenclature system for *S. enterica* based on DNA comparison studies linked to serological analysis. The complete description of a *Salmonella* now uses the following typical designation: *S. enterica* subspecies *enterica* serotype Typhimurium (or Dublin, Gallinarum etc. depending on the serovar) and this can be shorten to *Salmonella typhimurium* or now more commonly *Salmonella* Typhimurium [38].

The bacteria of the genus *Salmonella* are commonly classified using a serological scheme developed by Kauffmann and White (known accordingly as the Kauffmann-White Scheme) [39]. Indeed, this is the approach used by most current reference microbiology laboratories. The scheme works by raising typing sera in rabbits against key antigens present on the surfaces of the *Salmonella*. These typing sera work as references. The typing antigens are diverse and facilitate discrimination and groupings simultaneously. O antigen, a component of bacterial lipopolysaccharide, is one of the key typing antigens. There is significant diversification in the somatic O typing antigen (repeat units of saccharide, which give the smooth appearance to colonies growing on agar). Another key typing antigen is the flagellin protein or H antigen (a heat labile antigen located in bacteria flagellae). H typing can be based on *Salmonella* phase 1 and phase 2 flagella antigens [39]. A third class of antigen is the surface component known as "Vi" (a capsular polysaccharide that contributes to the virulence of the bacterium in the host). Vi is a linear, acidic homopolymer of α-1,4-

linked *N*-acetylgalactosaminuronate (D-GalNAcA), variably O-acetylated at C-3 [40, 41] and is commonly found on isolates of *S.* Typhi, *S.* Paratyphi C and occasionally on isolates of *S.* Dublin.

The Kauffmann-White scheme is significantly discriminating and has proved to be a robust, if not completely accurate approach to typing *Salmonella*. There are currently over 2400 reported *Salmonella* serovars and others are likely to emerge in the future. The frequency of isolation of *Salmonella* of a particular serovar varies over time and location with evidence of frequent epidemics and outbreaks. *S.* Typhimurium, *S.* Enteritidis and *S.* Dublin are commonly isolated serovars associated with human gastroenteritis.

S. enterica can cause a range of different disease syndromes. Certain serovars are regarded as more likely to cause gastroenteritis in a particular host (for example, *S.* Enteritidis in humans), whereas others are more likely to cause systemic typhoid (for example, *S.* Typhi in humans). This is a useful, but not an absolute classification, and disease outcome can be influenced by a range of factors such as the host/isolate pairing, immune status and infectious dose. Nevertheless, *Salmonella* are often classified based on the most common clinical outcome and in this context are described as either typhoidal (typhoid fever and paratyphoid fever) or non-typhoidal *Salmonella*.

1.2.1 Typhoidal *Salmonella*

Originally isolated in 1880 by Karl J. Eberth, the causative organism of typhoid fever *S.* Typhi is a pathogen that can colonise the lymphatic tissues of the small intestine, liver, spleen, and bloodstream of infected humans. It does not cause disease in animals, other than higher primates under experimental conditions [42]. Most *S.* Typhi isolates from typhoid fever cases express the polysaccharide capsule Vi, which is associated with increased infectiousness and virulence [43]. Humans are the only natural host and reservoir for *S.* Typhi although the pathogen can survive for days in water and for months in contaminated food. Reported risk factors include a history of contact with other patients before illness, access to dirty water and past evidence of infection with *Helicobacter pylori* [44]. Work on typhoid fever patients in Vietnam

has suggested an important role of HLA-linked genes in governing susceptibility or resistance to this infection. HLADRB1* 0301/6/8, HLA-DQB1*0201-3, and the tumour necrosis factor A loci (TNFA*2-308) were associated with susceptibility to typhoid fever, while HLA-DRQB1*04, HLA-DQB1*0401/2 and TNFA*1(-308) were associated with lower risk [45, 46]. HLA-DRB1*12 is associated with protection against complicated typhoid fever [47].

The clinical presentation of typhoid fever is variable, ranging from fever with little morbidity to marked toxaemia and associated complications involving many systems. The commonest complications are gastrointestinal bleeding, intestinal perforation and typhoid encephalopathy [48, 49]. In endemic regions, diagnosis can be missed because of non-specific features like diarrhoea and vomiting, or predominant respiratory symptoms. In children younger than 5 years, typhoid fever can be milder and can mimic a viral syndrome. The rate of severe complications is lower than at later ages. Factors affecting severity could include duration of illness before therapy, choice of antimicrobial therapy, strain virulence, inoculum size, previous exposure or vaccination, and other host factors such as HLA type (see above), immune suppression or antacid consumption.

Clinical features of paratyphoid fever are generally reported to be similar to those of typhoid fever but are usually thought to be milder with a shorter incubation period. *S.* Paratyphi A or Paratyphi B can manifest with jaundice, thrombosis, and systemic infections. *S.* Paratyphi B might occasionally have an onset similar to non-specific gastroenteritis [50]. Gastrointestinal symptoms are usually not present with *S.* Paratyphi C but there have been cases with systemic complications such as septicemia and arthritis [51].

Unlike many *Salmonella* infections, typhoid can be associated with a chronic, potentially asymptomatic carrier state involving systemic tissues such as the gallbladder. Chronic typhoid carriers are likely important for survival of the pathogen, and may be responsible of the contamination of water and food. Carriers can be notoriously difficult to identify because they are usually quite healthy, although elevated levels of anti-Vi antibodies can be present in the serum of carriers [52].

1.2.2 Nontyphoidal *Salmonella* (NTS)

The NTS are more frequently associated with localised acute gastroenteritis and diarrhoeal disease rather than typhoid. Acute gastroenteritis is the most common presentation of NTS infection. Typical symptoms include diarrhoea, nausea, headache, and sometimes vomiting. Fever and abdominal cramps are almost always present. Bloody diarrhoea and invasive disease may occur, particularly with certain serotypes. NTS can cause invasive disease, particularly in compromised hosts. Invasive infection may present as urinary tract infection, septicaemia, abscess, arthritis, cholecystitis and rarely as endocarditis, pericarditis, meningitis, or pneumonia. Asymptomatic carriage can occur in as many as 5% of healthy hosts [53].

NTS bacteria are widely distributed in the animal kingdom, including humans, livestock, pets, wild mammals, poultry (and other birds), reptiles and amphibians as well as in seafood. Most NTS serovars are regarded as being more promiscuous in terms of their abilities to infect different hosts, compared to the typhoidal serovars that are often host-adapted or even host-restricted. As NTS are generally promiscuous they often have zoonotic potential, surviving in veterinary herds or companion animals, from which they can spread to humans via food consumption or environmental contamination. Although NTS may cause disease in one animal they may just colonise other species or older members of the same species. Thus, the status quo of the host/pathogen relationship can vary. Indeed, NTS that frequently cause gastroenteritis in humans, such as *S*. Typhimurium and *S*. Enteritidis, can cause invasive disease in the compromised [54-56]. Approximately 2 to 8 percent of NTS infections are associated with bacteraemia, and are not always preceded by gastroenteritis. Risk factors for NTS bacteraemia include being immunocompromised (including HIV, malignancy, chemotherapy, steroid therapy) and extremes of age (less than 3 month and greater than 50 years old). However, such risk factors are not apparent in up to one third of cases of NTS bacteraemia. Extra-intestinal focal infections such as arthritis, meningitis or pneumonia occur in 5-10% of those with bacteraemia. Additionally, NTS can be more associated with invasive disease in particular settings, for example in sub-Saharan Africa where the epidemiology may be different and the genotype of the isolates may be specialised

[16]. Further, there may be significant differences on pathogenic potential even within a serovar. For example, DT2 isolates of *S.* Typhimurium are largely host restricted to pigeons and do not efficiently infect humans [57]. Common transmission routes for NTS include ingestion of contaminated water or food, direct exposure to infected animal or their waste as well as faecal-oral transmission. Infection can also occur in medical care settings where immunocompromised patients are at increased risk.

1.3 Molecular approaches to identify and discriminate between *S. enterica*

The ability to identify and discriminate between microbes associated with disease is important for epidemiological surveillance and facilitating public health policy decision-making. Many different methods have been developed to identify pathogens with relatively varying degrees of success. Microbial culture is the classical approach, although not all microbes can be readily cultured. Once, cultured, microbes can be subjected to different phenotypic and molecular tests. If the microbe cannot be cultured, then sensitive molecular or immunological assays may be more appropriate. The most common traditional phenotypic assays applied to cultured microbes include the use of serological and metabolic tests, although such assays may be challenging in terms of their specificity and discriminatory power. Multi-Locus Enzyme Electrophoresis (MLEE) is a metabolic assay that measures different enzyme activities and this approach proved useful in the early days of defining *S. enterica* phylogeny [58]. The Kauffmann-White scheme has proved to be a robust methodology for serotyping *S. enterica* isolates, with over 2000 different serovars being defined to date. These serovars can be allocated to generalised typhoidal and non-typhoidal classes and into host adapted or host restricted types (Figure 1.3).

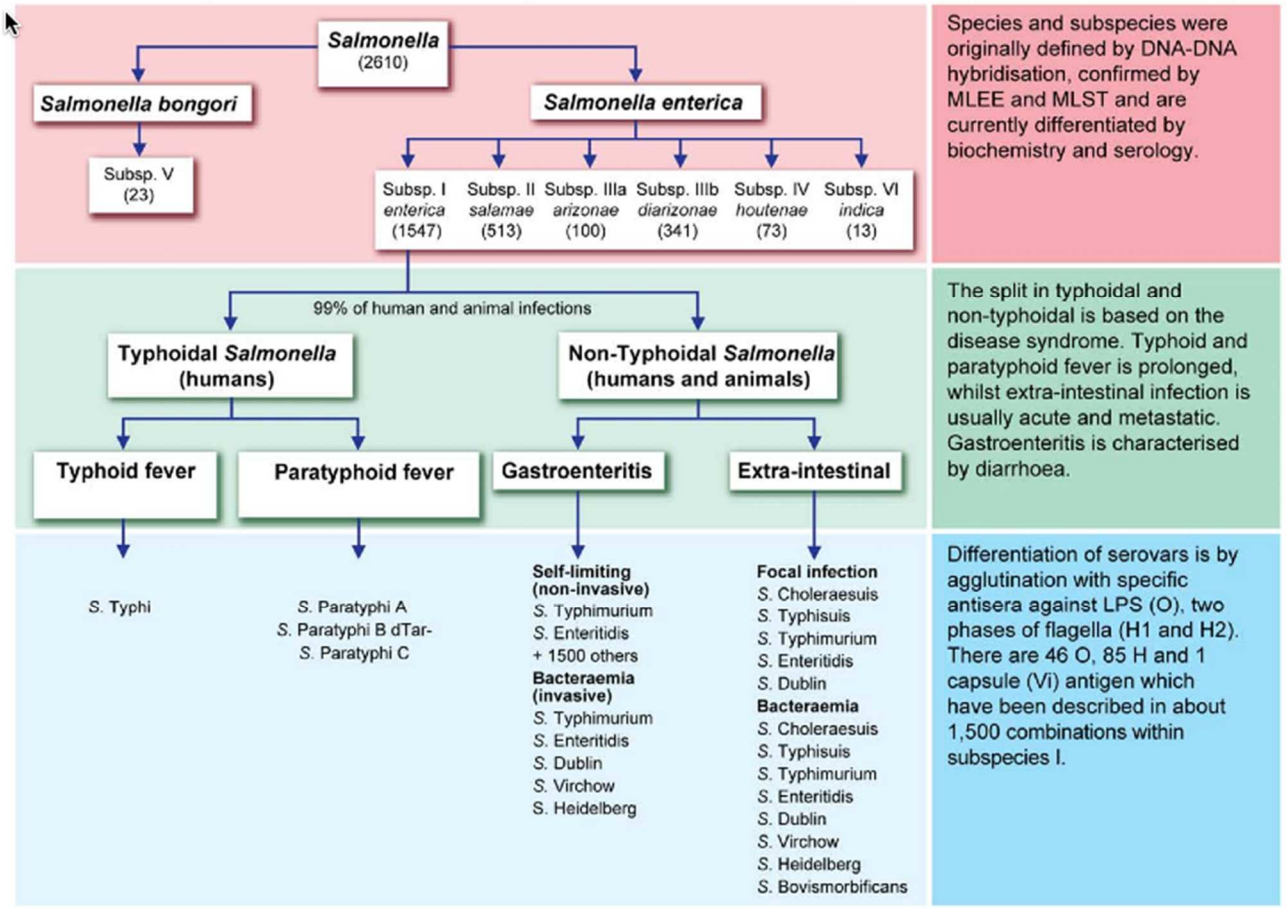


Figure 1.3: General overview of the current classification of *S. enterica.* Taken from [59]

Serotyping in *S. enterica* relies on specific reference sera that are generated by immunising rabbits or other animals. These reference sera are difficult to quality control and are usually generated in reference laboratories or by companies. Thus, serotyping can be challenging and expensive and is utilised predominantly in reference laboratories. Nevertheless, the Kauffmann-White scheme has proved to be of tremendous value in terms of defining *S. enterica* serovars over decades, facilitating diagnosis and outbreak analysis. However, serotyping does not readily discriminate below the serovar level and provides little or no phylogenetic information.

Serotyping relies on detecting antigenic variation in a limited number of surface associated bacterial antigens. Another target for the identification and typing of microbes is the genome. Analysis targeting DNA can indirectly measure differences in genome organisation (chromosome, plasmids etc.) or can be directly DNA sequence based. Techniques that analyse genome organisation include plasmid profiling [60] and Pulsed Field Gel Electrophoresis (PFGE) [61]. Plasmid profiling has limited utility as plasmids represent only a small component of the genome and can transfer between isolates, compromising identification. PFGE works by targeting rare restriction sites on the chromosome and analysing large DNA fragments generated by restriction using pulse field electrophoresis. This approach has found broad utility in the area of *S. enterica* molecular epidemiology and has proved extremely useful for analysing global spread and local outbreaks [62, 63]. Specific software has been developed to image and compare the patterns of DNA fragments generated by PFGE and networks such as PulseNet have appeared to facilitate data exchange. Nevertheless, PFGE provides only limited genomic and phylogenetic information and lacks discriminatory power and methods based on whole genome analysis are likely to supersede them. Other sub-genomic methodologies have also been developed based on Polymerase Chain Reaction (PCR) methods, including Restriction Fragment Length Polymorphism (RFLP) [64]. These have proved useful for discrimination, but are also likely to be superseded by whole genome sequencing.

DNA sequence based methodologies have the advantage of simplifying comparative analysis. Such typing methods exploit the unambiguous nature and electronic portability of nucleotide sequence data to classify microorganisms. Multi-Locus

Sequence Typing (MLST) was one of the first methodologies developed that was based on sequence reads and the approach has found broad utility for different pathogens, including *S. enterica* [59, 65]. Classical MLST is based on sequences derived from 7 housekeeping gene fragments (alleles) that represent the core genome of a microbe. The 7 sequences generated by different isolates can be compared at the nucleotide level and differences quantified. Isolates that possess identical alleles for all gene fragments are assigned to a common Sequence Type (ST), and STs that share all but 1 or 2 alleles can be grouped into ST-based clonal complexes on the basis of software such as eBurst [66, 67]. This scheme has also been used to survey the genetic properties of various *S. enterica* serovars, including antibiotic resistant clades. The results suggested that most of the time, the MLST type correlates with serovar, with some exceptions [59]. eBurst analysis provides some phylogenetic information and allows sub-serovar discrimination within many serovars. A recent study using MLST of the population structure of subspecies in *S. enterica* showed that many *Salmonella* STs cluster together in discrete groups called eBGs (eBurstGroups) [59]. Here, an eBG was defined by groups of 2 or more STs that were connected by pair-wise identity at 6 of the 7 gene fragments, thus sharing 6 of the 7 alleles that defined the ST. However, some serovars, such as *S.* Typhi, are significantly monophyletic and here the utility of MLST is somewhat limited. Nevertheless, it has been proposed that MLST analysis, or similar approaches, could replace serotyping. Variants of MLST have also been developed that target specific genetic loci (for example, O antigen loci) or genes (for example, *fliC* for flagellin) [68].

In general, bacterial taxonomy can exploit a top-down approach based on phylogenetics in order to elucidate a genealogical tree or a bottom up analysis exploiting population genetics in order to identify populations and networks. Trees are appropriate for clonal organisms. Population genetic analysis is arguably more appropriate for organisms with frequent homologous recombination such as *S. pneumoniae* [69]. *S. enterica* does exhibit recombination but not at the level of the more recombinogenic species. In reality, bacterial taxonomy benefits generally from a combined approach. Indeed, new approaches are in development, which explicitly include lateral gene transfer events in the genealogy. Considering that frequent recombination has the potential to alter classical phylogenetic data, definition of

eBurst groups based on allelic identity rather than sequence identity provide discrete clusters of related organisms even in presence of significant levels of homologous recombination [59].

1.4 Evolution of the genus *Salmonella*

The genus *Salmonella* is thought to have evolved from a common ancestor within the *Enterobacteriaceae* over many millions years, with estimates of at least 100 million years of such evolution [70]. The two recognised *Salmonella* species *S. bongori* and *S. enterica* are thought to have diverged 40 - 65 million years ago although such lengths of time are very difficult to estimate. The evolutionary signatures that mark differentiation and those that track different branch points have been investigated by analysing factors including sequence variation, gene flux across the species/subspecies, the distribution of the different virulence-associated systems and metabolic traits.

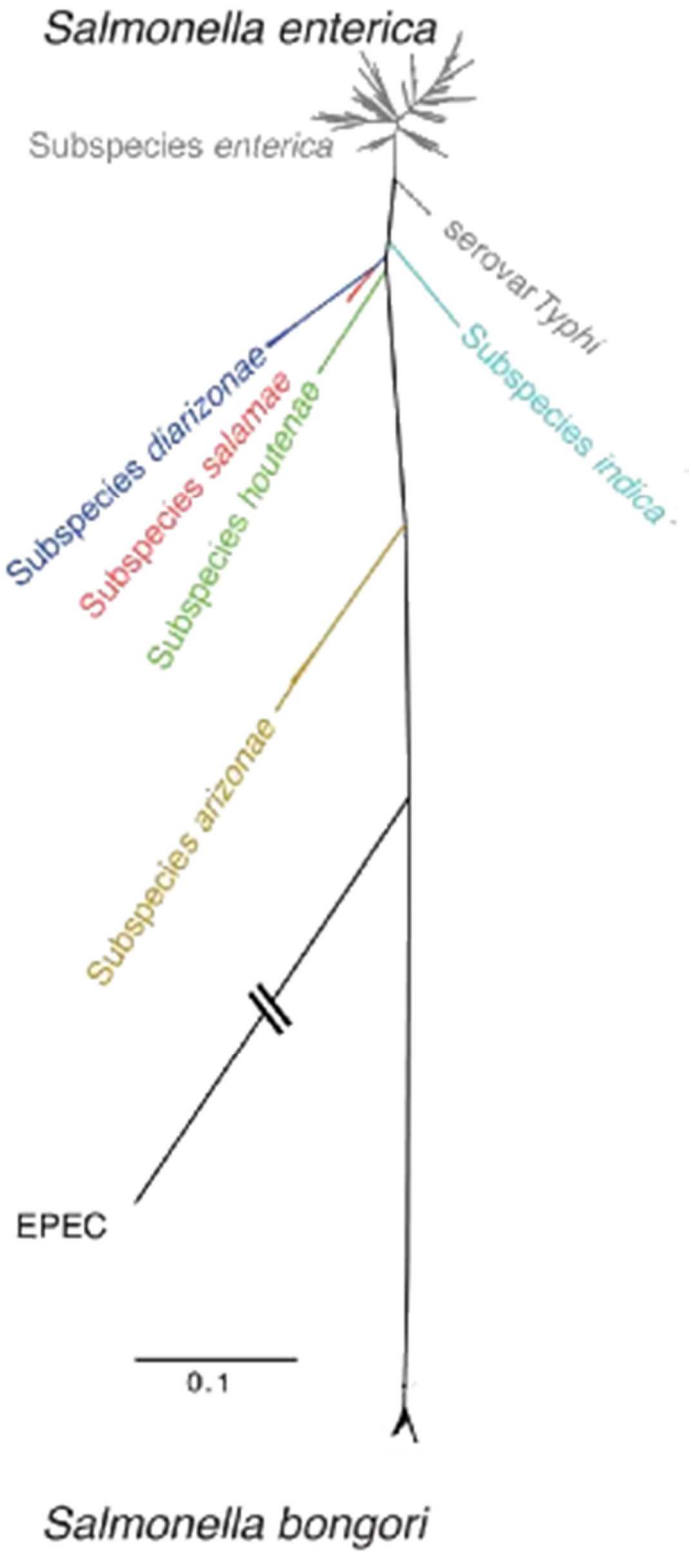


Figure 1.4: Maximum likelihood phylogenetic tree of *Salmonella* based on concatenated MLST loci. Taken from [71]

A combination of MLST and whole genome analysis (see Section 1.3) can be used to provide an overview of the general organisation of the *Salmonella*. For example, Figure 1.4 displays a candidate phylogeny based on concatenated MLST analysis. Such analysis places *S. bongori* on a separate evolutionary branch from *S. enterica*, suggesting a common ancestor but no direct succession [71]. *S. bongori* is likely the older species that evolved largely, and remains predominantly, within cold blooded animals, including reptiles. The various *S. enterica* sub-species can be seen distributed along the phylogeny leading to sub-species I, which has undergone significant relatively recent expansion, likely within predominantly warm blooded animals, including humans.

Gene flux and exchange can provide insight into the evolution of the *Salmonella*. For example, the analysis of the presence or absence of major virulence-associated systems is a useful approach to trace the evolution of *S. bongori* and *S enterica*. There are currently 22-reported Salmonella Pathogenicity Island (SPIs) and among those, only SPI-1, SPI-4 and SPI-9 are present in the reference genome *S. bongori* 12419. These shared islands have a similar gene composition in each species, although there is significant sequence drift. However, the other *S. enterica* SPIs are either incomplete or absent. This indicates that there are significant differences in the pathogenic potential of the two species in different hosts. An important distinguishing feature of *S. bongori* is the absence of SPI- 2. The site occupied by SPI-2 at tRNA–*valV* in *S. enterica* is occupied by a 20 kb genomic island in *S. bongori* encoding a novel type VI secretion system called SPI-22, although the tetrathionate respiration (*ttr*) gene cluster present in SPI-2 is retained by *S. bongori*.

Although *S. bongori* lacks the 4 distinct T6SSs described for *Salmonella*, encoded on SPI-6, SPI-19, SPI-20 and SPI-21 [71], the T6SS genes carried on SPI-22 share extensive similarity with the T6SS of other *Enterobacteriaceae*. These include the recently identified CTS2 T6SS locus of *Citrobacter rodentium* ICC168 [72] and the HSI-III locus of *Pseudomonas aerugin*osa PA01 [73]. Thus, in this regard *S. bongori* broadly resembles the wider *Enterobacteriaceae* or ancestral state. In addition to the lack of SPI-2, *S. bongori* lacks part of SPI-6 (encoding a type VI secretion system), SPI-13 (necessary for survival in chicken macrophages), SPI-14 (involved in electron transport system) and SPI-16 (a bacteriophage remnant, carrying genes

associated with LPS modification). These and other potential virulence-associated factors absent from *S. bongori* may go some way to explaining the limited ability to cause disease in warm-blooded animals demonstrated by the introduction of SPI-2 into *S. bongori* [74]

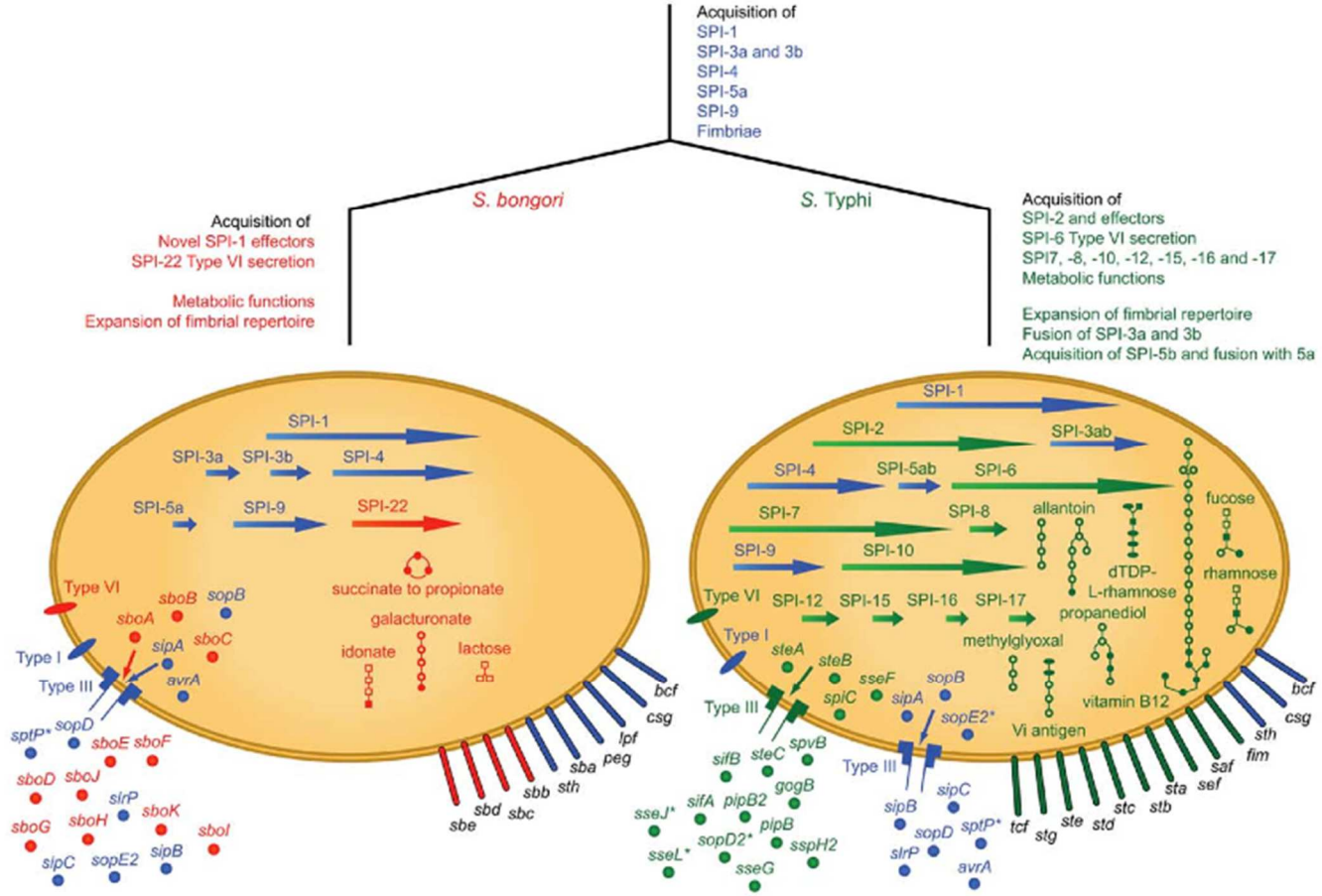


Figure 1.5: Diagram summarising selected aspects of the evolutionary history of *S. bongori* and *S.* Typhi, a comparative member of *S. enterica.* Taken from [71]

Metabolic functions have also been shown to exhibit significant evolutionary divergence between *S. bongori* and *S. enterica*. For example, *S. bongori* possess only fragments of the first and last gene of the *cob-pdu* gene cluster and therefore lacks the faculty to anaerobically synthesise vitamin B12 and to catabolise propanediol [74]. The *cob-pdu* gene cluster is also absent/lost from most *Enterobacteriaceae* and may have been independently acquired by *S. enterica* [75, 76] potentially as an adaptation to the tissues and cells (macrophages) of warm blooded animals [76], a niche where *S. bongori* is arguably poorly adapted [74]. Interestingly, some host restricted *S. enterica*, including *S.* Typhi, harbours inactivating mutations (pseudogenes) in the *cob-pdu* system and this may, in some way, facilitate host adaptation.

Another example of metabolic differences includes the ability to ferment L-tartrate and citrate, which are postulated by others to differentiate high and low pathogenicity *Salmonella* strains [77]. Additionally, in *K. pneumoniae*, the ability to ferment citrate partially divides clinical isolates into two groups and may represent an adaptation to different nutrient conditions found within the host [78]. Interestingly *S. bongori* is able to utilise lactose and like *E. coli* is lactose-positive. In *E. coli*, the lactose system may facilitate the metabolism of milk sugar and adaptation to the mammalian gut but it is not clear what the advantage, if any, is for *S. bongori*. In contrast *S. enterica* are lactose-negative and it has been proposed that *lacI* expression can interfere with the function of SPI-2 and attenuates virulence in macrophage [79].

1.5 Major virulence-associated genes

Salmonella are well adapted to their lifestyle of colonising their hosts through the intestinal and transmitting between hosts via the environment. They also are highly invasive in terms of their ability to enter both the local tissues associated with the intestine and the cells that make up these tissues. The ability to invade both tissues and cells is central to the pathogenesis of *Salmonella* infections and is fundamental for their ability to survive within their hosts and spread between them, Clearly *Salmonella* differ significantly in aspects of their pathogenicity, including their ability to invade their host tissues. Some of this has already been discussed in relation to typhoid and nontyphoidal isolates.

S. enterica harbour multiple genes that contribute to virulence and survival. Screens involving high throughput mutagenesis and gene tagging technologies such as Sequence Tagged Mutagenesis (STM) and Transposon Directed Insertion site Sequencing (TraDIS) have identified scores of virulence-associated genes in different *Salmonella* isolates [80]. *Salmonella* bacteria are believed to target a variety of phagocytic and non-phagocytic cells in vivo. Following ingestion, *Salmonella* can be either passively (opsonisation, phagocytosis) or actively (promoted by the bacteria) internalised within the host cells through different processes. One of the most common internalisation mechanisms is phagocytic uptake, particularly by monocytes or macrophages. The phagocytosis of *Salmonella*, and indeed other bacteria, is a complex series of steps involving multiple receptors and potentially antibodies and complement factors. Once phagocytosed, pattern-recognition receptors on or within the phagocytic cell can recognise pathogen-associated molecular patterns such as LPS or flagellin, which bind to their respective ligand, either on the cell surface or inside the *Salmonella*-associated phagosome [81].

Alternatively, *Salmonella* can also actively target both phagocytic and non-phagocytic cells using a type III secretion system (T3SS), designated T3SS1 or Salmonella Pathogenicity Island 1 (SPI-1). T3SS1-mediated invasion by *Salmonella* is a very specific process depending on a highly regulated expression of a number of factors that mediate invasion [82, 83]. The T3SS1 apparatus can be visualised as a so called needle like complex that facilitates contact with this host cell and secretion of effector proteins that prepare the targeted host cell for invasion. The genes encoding the expression of the T3SS1 apparatus are largely encoded within the SPI-1 locus [84]. A number of effector proteins (for example, SipA, SipC, SopB/SigD, SopD, SopE2 and SptP) are actively secreted through the T3SS1 needle into the host cell. There they act in a coordinated manner to induce dramatic rearrangement of the actin cytoskeleton resulting in membrane ruffling and rapid internalisation of the bacteria in a highly engineered process [85-87]. The exact complement of T3SS1 effectors can vary between different Salmonella serovars. However, the T3SS1 system is present in *S. bongori* and *S. enterica* and is a fundamental virulence-associated factor of the *Salmonella*.

Of the effector proteins translocated by T3SS1 into the host cell, SopE/SopE2, SopB and SipA, are known to play a role in inducing the actin rearrangements required for invasion and ruffling [86, 87]. Most of the other effector proteins are implicated in multiple post-invasion processes such as modifying host cell survival, forming the Salmonella Containing Vacuole (SCV) and modulation of the inflammatory response. For instance, some evidence suggest that the inositol phosphatase SopB plays a role in Akt activation, fluid secretion and SCV formation during the invasion process [88-90]. SopB is localised within the SCV and its activity may be influenced by ubiquitination. Another T3SS1 effector AvrA, an ubiquitin-like acetyltransferases/cysteine protease, removes ubiquitin from IkBa and beta-catenin, 2 inhibitors of the NF-kB pathway, thereby inhibiting the inflammatory response [91], activating beta-catenin signalling [92, 93] and preventing apoptosis in intestinal epithelial cells [91]. SopA, another invasion modulator, harbours a HECT-like E3 ubiquitin-ligase activity [94]. The tyrosine phosphatase SptP another T3SS1 effector involved in SCV formation is required for switching off ruffle formation following invasion. SipA has been shown to influence SCV morphology.

Fimbriae, non-fimbrial adhesins and flagella on the surface of *Salmonella* may also mediate bacterial attachment and consequently contribute to internalisation via processes independent of phagocytosis or T3SS1-mediated invasion [95]. Again, the repertoire of fimbria and other adhesins show significant variation within *S. enterica*, as do the type of flagella so the mechanisms of attachment may vary significantly. *S.* Typhimurium and other serovars harbour multiple fimbrial loci, many of which are only induced *in vivo*. Fimbriae can have a role in biofilm formation, attachment to host cells and colonisation of the intestine [96].

Motility has been associated with the invasiveness of *Salmonella* [97]. Indeed, different flagella types may significantly influence the attachment, invasion and activation of cells targeted by *Salmonella*, for example, in *S.* Typhi. Within macrophages, flagellin can be translocated into the cytosol by the T3SS1, resulting in activation of the inflammasome and caspase-1-mediated cell death (pyroptosis) [98, 99]. In the intestinal epithelium flagellin induces inflammation while inhibiting apoptosis via TLR5 in basolateral epithelial cells. Flagella are generally down-

regulated inside the host except in macrophages potentially limiting recognition by TLR5.

Once internalised into the host, bacteria that successfully reach the sub-mucosa need to survive and replicate within different host cells in order to establish a sustainable infection. Dendritic cells may play a role in the initial penetration of bacteria across the mucosal surface involving a mechanism by which dendrites reach through the epithelial barrier and engulf *Salmonella* [100]. Activated dendritic cells may thus be a portal of entry into host tissues for the invading *Salmonella*. However, once within the tissues *Salmonella* are believed to rapidly enter macrophages by largely unknown mechanisms. Indeed, multiple data has highlighted a central role for macrophages in the survival and persistence of *Salmonella in-vivo* [101].

Non-fimbrial proteins have been associated with enhanced *Salmonella* adhesion and subsequent invasion into cells. These include BapA and SiiE; 2 surface-associated proteins secreted via the type I secretion systems BapBCD and SiiCDF respectively. SiiE is encoded on SPI-4, which is co-regulated with SPI-1 [102]. Other proteins including RatB, SivH and ShdA encoded on the CS54 pathogenicity island have been shown to contribute to the persistence and shedding of *Salmonella* in the intestine by targeted connective tissue proteins such as fibronectin [103].

T3SS2 or Salmonella Pathogenicity Island 2 (SPI-2) contributes to systemic virulence and survival/persistence within macrophages. Although the different roles of individual T3SS2 effectors are not fully characterised, some of them have been associated with SCV formation and positioning within the cell. For instance, SseF and SseG are required for maintenance of the SCV and facilitating intracellular replication [104-106]. SifA plays a role in SIF (Salmonella Induced Filaments, visible by microscopy) formation, a process linked to maintaining SCV membrane integrity [107]. Others T3SS2 effectors, such as PipB2 and SseJ, cooperate with SifA, further influencing SCV membrane integrity; PipB2 interacts with kinesin light chain, a subunit of the kinesin-1 motor complex by recruiting it to the surface of the SCV [108] while SseJ, promotes host membrane tubulation [109]. In epithelial cells infected with mutants lacking SseJ cholesterol accumulation is increased compared with cells infected with wild-type bacteria, and this is associated with a decrease in

intracellular replication. Here, it is worth noting that SseJ is a pseudogene in *S.* Typhi. SseL, can also modulate NF-kB activation downstream of IkBa kinases although its specific role remain unclear. Again different *S. enterica* can harbour different combinations of SPI-2 associated effectors.

1.5.1 Virulence-associated plasmid

Isolates from a number of different serovars of *S. enterica* harbour a plasmid associated with systemic virulence in the mouse, known as the Salmonella Virulence Plasmid. This plasmid is absent in *S.* Typhi and *S.* Paratyphi A. The *spvRABCD* genes are located on these plasmids and are key plasmid-mediated virulence-associated factors in some serovars. The virulence plasmid is present in many isolates of *S.* Typhimurium and other gastroenteritis-associated serovars such as *S.* Enteritidis [110, 111]. SpvB and C may be translocated into the host cells via T3SS2 or plasmid encoded genes. SpvB ADP-ribosylates actin destabilises the cytoskeleton and is associated with host cell cytotoxicity [112].

1.5.2 Examples of other virulence-associated genes

Many other genes have been associated with *Salmonella* virulence in different models and hosts and some of these will be considered here. For a more detailed analysis of *Salmonella* virulence-associated determinants please consider these reviews and screens [113-115]. Many host phagocytic cells produce reactive oxygen species through the phagosomal NADPH oxidase (NOX2) complex as a defence mechanism for killing intracellular pathogens. To counteract this activity, *Salmonella* can express superoxide dismutases such as SodCI for protection against extracellular reactive oxygen species. SodCI is tethered within the periplasm of the phagosome and is significantly protease resistant [116]. The level of iron in host tissues and cells is tightly regulated to control direct access for pathogens. In the host, free iron binds to iron-binding proteins such as transferrin where it is largely unavailable to bacteria without specialised acquisition systems. The host has other mechanisms to deny iron to pathogens such as *Salmonella*. For example, Nramp1 is a divalent metal-proton transporter found in key protective cells such as macrophages, neutrophils and dendritic cells [117] that creates a restricted availability of free iron for the bacteria

by limiting iron in the phagosome. This effectively limits the ability of bacteria and other pathogens to establish an infection. In response to iron deprivation, *Salmonella* produce siderophores, including enterobactin and salmochelin [118]. Salmochelin is a glucosylated derivative of enterobactin and this modification may be important for resistance to lipocalin-2, an antimicrobial protein that prevents bacterial iron acquisition in the inflamed intestinal epithelium [119]. A recent study using different *S.* Typhimurium mutants lacking iron transporters has shown that iron transporters encoded by *feoB* and *sitABCD* are required for optimal survival in Nramp1 -/- mice and replication in macrophages. Additionally, the Nramp1 homologue MntH, which prefers Mn(II) over Fe(II), is also required for optimal virulence [120].

Salmonella has acquired various proteins for the uptake of Magnesium including CorA, MgtA and MgtB, which are essential for virulence in different models [121]. Mg^{2+} transporters are required for intra-macrophage survival and growth in magnesium-depleted medium. K^{+} and Zn^{2+} are also implicated in intracellular survival; *ZnuABC S.* Typhimurium mutant derivatives are defective for virulence in both susceptible and resistant mouse strains [122]; ZnuABC is a high-affinity Zn^{2+} transporter in low- zinc conditions. The Trk system functions as a low-affinity K^{+} transporter and may be involved in resistance to antimicrobial peptides [123].

1.6 Signatures of adaptation in *S. enterica*

Most *S. enterica* serovars are classically associated with a broad host range. However, a few serovars are significantly host-restricted. For example, *S.* Typhi and *S.* Paratyphi A are highly human adapted, whereas *S.* Gallinarum isolates are poultry (bird) adapted. Even within serovars such as *S.* Typhimurium there is evidence of isolates or clades being adapted or restricted to particular hosts. For example, within *S.* Typhimurium DT2 phage type isolates display avian (for example, pigeon) adaptation and ST313 isolate may be adapted to humans [124].

In recent years, comparative genomic studies of broad-host-range serovars, which are believed to be the ancestor-state of host-restricted serotypes, have provided insights into the genomic signature of bacterial host adaptation and evolution. These studies have identified a number of genome signatures that may represent evidence of host

adaptation leading to restriction. These adaptive/restrictive signatures include both gene acquisition and inactivation.

Horizontal gene transfer has been broadly implicated in the evolution of virulence and indeed resistance to antimicrobials through plasmids and transposons. In bacteria, horizontal gene transfer is recognised as a general mechanism driving evolution. The key role of mobile genetic elements in the acquisition of virulence traits in bacteria has been extensively studied and several reports have associated mobile virulence-associated determinants with host adaptation or the occupation of a new niche within the host [125]. Within the *E. coli*, horizontal transfer events have facilitated the transition from a commensal to a pathogenic lifestyle. Examples include the acquisition of the heat-labile and stable toxin genes on plasmids and the LEE (Locus for Enteric Effacement) and intimin genes of EPEC lineages [126].

In *Salmonella*, the evolution of virulence has been driven by the incorporation of distinct genetic elements into the genome including pathogenicity islands, T3SSs and the *Salmonella* virulence plasmid. The virulence genes acquired by horizontal transfer have to be incorporated into existing gene expression regulatory circuits to ensure coordinate expression of virulence-associated genes in a manner that does not compromise fitness and competitiveness. Horizontal gene transfer in *Salmonella* can involve phages which are frequently exchanged even within clades or serovars [127]. Indeed, *S.* Typhimurium phages Gifsy-1, Gifsy-2 and Gifsy-3 have successfully lysogenised a range of serovars and different lineages of *S.* Typhimurium [128].

Within *S.* Typhimurium a potentially mobilisable element known as SPI-7 is associated with the acquisition of the locus encoding Vi capsule that is directly linked to virulence in humans and hence host adaptation [129]. Vi makes *S.* Typhi more resistant to antibody directed killing and complement mediated phagocytosis. Vi is also immunomodulatory and may facilitate the ability of *S.* Typhi to invade tissue without inducing inflammation, potentially by enhancing interleukin 10 production [130]. SPI-7 also encodes a SopE phage and a Type IV pilin that have also been associated with virulence. Hence, the acquisition of SPI-7 is a clear example of horizontal gene transfer influencing host adaptation. *S.* Typhi and *S.* Paratyphi A have also recently been shown to encode a novel toxin named typhoid

toxin that may also influence human infectivity [131]. The distribution of the genes encoding this toxin is limited only a few *S. enterica* serovars.

Within *Salmonella* lateral gene transfer through conjugation can also transfer antibiotic resistance determinants [132]. Acquisition of plasmids encoding resistance and virulence properties has been known to influence bacterial evolution. Thus, plasmids can be acquired by horizon gene transfer from other serovars, or even other species or genuses. For example, *S.* Typhi CT18 possesses a 218,150-bp multiple-drug-resistance incH1 plasmid (pHCM1) and a 106,516-bp cryptic plasmid (pHCM2), which shows recent common ancestry with a virulence plasmid of *Yersinia pestis* [133]. In *S. enterica,* many high molecular weight plasmids encode virulence-associated genes or are responsible for antibiotic resistance. As discussed, the classical *Salmonella* virulence plasmid encodes the *spvRABCD* genes involved in intra-macrophage survival of *Salmonella* but this plasmid has also been shown to be able to acquire antibiotic resistance genes that have the potential to spread in bacterial populations [16]. Many low molecular weight plasmids have been found in *S. enterica* but in general little is known about their function, although some studies have suggested a role in increasing resistance to phage infection due to the presence of restriction modification systems [134]. Despite limited knowledge on their function, their presence or absence is frequently used for strain differentiation in epidemiological studies.

There is now extensive evidence, gathered from genetic and genomic analysis, that bacteriophages are drivers of evolution in the enteric bacteria, including within *S. enterica*. Many intestinal commensals and enteric pathogens harbour prophages or phage remnants integrated within their genomes, often at multiple sites [128]. Additionally, prophages can encode so called ‘cargo’ genes that are not required for phage growth but can encode virulence-associated factors that can influence pathogenicity [135]. The diversity of prophage within a bacterial population is influenced by transduction and recombination involving superinfecting phages, resident prophages, or occasional acquisition of other mobile DNA elements. Prophages also play a part in the diversification of the genome architecture and represent strategic points for genome insertions and inversions [135].

1.6.1 Genome degradation/decay

Genome decay or reductive evolution is a process by which bacteria lose some functions by gene deletion or degradation (for example, the acquisition of frame shifts or stop codons). Such potentially inactivated genes are often referred to as so called pseudogenes. This process has been reported in obligate intracellular parasites, such as *Rickettsia prowazekii, Mycobacterium leprae* and *Chlamydia* spp. [136-138] as well as in *Yersinia pestis* [139]. As genes become inactivated they may restrict bacteria to specialist hosts or novel niches within their hosts. Thus genome degradation is a signature of host restriction and niche change. In *Y. pestis* and *S.* Typhi the host is humans and the niche change is from the intestine into the systemic system. Extensive genome degradation has been observed in *R. prowazekii*, the typhus agent, with only 76% of the potential coding genes being likely fully functional [140]. Another example of genome reduction was documented in *M. leprae*, which may originally have had a genome similar in size to other Mycobacteria (around 4.4 Mb) but this has been downsized during evolution through rearrangement. *M. leprae* may have lost more than 2,000 genes [136] and this might explain its extremely slow replication rate, lack of acute disease and targeting of neurones.

Genome degradation has also been documented in *S.* Paratyphi A and *S.* Typhi where both share components of their genomes and have similar phenotypes (human-restricted and systemic disease). Around 170 pseudogenes are present in *S.* Paratyphi A, whereas *S.* Typhi can harbour over 200 [141]. Several of these pseudogenes correspond to genes known to contribute to virulence in *S.* Typhimurium and other more promiscuous *Salmonella* serovars. About 30 genes are degraded in both *S.* Typhi and *S.* Paratyphi, although the inactivation of different genes in common pathways is more common (for example, in chemotaxis, vitamin B12 acquisition and in the production of fimbriae) [142]. Amongst these 30 genes, several genes, such as *sopA* and *shdA* are known to be important in gastroenteritis and diarrhoea, which is uncommon in infections associated with these serovars.

1.7 Methods for phenotyping

A central aim of biomedical research is to fully understand the mechanisms of human disease and develop new and improved therapies or diagnostics. In order to achieve this, different disease models have been developed, although many of these fail to faithfully recapitulate the human condition. For this reason, researchers exploit different models in a complementary way to build a more informed picture of the human condition.

1.7.1 *In-vitro* models

While whole animal models have been tremendously useful for the current understanding of many human infectious diseases, it can be difficult to identify critical cellular and molecular contributors to disease using in vivo models. In order to address the role of specific host genes involved in the early host-pathogens interactions, many cell lines, have been successfully used as powerful tools to understand the mechanistic of infection. Non-polarised cell lines, such as HeLa and Hep-2 cells, although quite different from polarised cells of the intestinal epithelium, have been used extensively to study the cellular basis of the host-pathogen interaction. These cell lines have been used to unravel the molecular mechanisms of actin re-organisation [143, 144] and the elucidation of the bacterial and host proteins that contribute to this [145]. In addition, non-polarised cells had been used to elucidate the process of T3SS-dependent protein translocation into host cells as well as functional analyses of the injected effectors [146]. Taking advantage of their ease to grow and manipulate, tissue culture cells have successfully been used to analyse the biochemical activity of effectors such as Tir, in *E. coli* [147].

Non-polarised cells are not appropriate to study intestinal barrier function or maintenance of the brush border. Therefore, specific cells lines that mimic intestinal cells polarisation have been generated, such as MDCK, Caco-2, T84, and HT29 cells. These cell lines provide *in-vitro* models for investigating how bacteria disrupt epithelial barrier function during infection [148, 149] and for identifying effectors specifically required in the process. As with the non-polarised cell lines, these cell

lines offer the advantages of convenience, rapid growth, uniformity and availability of genetic tools for insertion of mutations. In addition to interactions with enterocytes, enteric pathogens also interact with phagocytic cells and dendritic cells. The use of macrophage-like cell lines such as J774 and U937 to model these specific immune cells has been essential for the identification of different effector functions and for targeting independent aspects of the phagocytic function of mammalian macrophages [150].

More complex models such as the IVOC (*in-vitro* organ culture) have been developed. IVOC exploits freshly obtained human (or other species such as cattle) intestinal biopsies, which are kept in tissue culture media under oxygen to delay cell death. The advantage of IVOC is that the infected tissue is close to native live tissue. However, compared with other *in-vitro* models, the use of IVOC for experimental infection is technically challenging and requires coordination with a clinic to obtain fresh tissue. In addition to the technical challenges, the variability in sampling methodologies between donors can result in differences in experimental outcome. The IVOC system has been used successfully to study host specificity and tissue tropism, including Tir/intimin-dependent colonisation and lesion formation by EHEC [151]. Human IVOC has also been used to study *S*. Typhimurium [152].

1.7.2 *In-vivo* model: The mouse

It is arguable that biomedical research has benefited significantly from the use of animal models to understand the pathogenesis of disease at a whole organism level. Additional, insight can be gained into biology at the cellular and molecular level. *In-vivo* models also provide systems for developing and testing new therapies in a preclinical setting. Mammalian models, such as the mouse, have been pre-eminent in modelling human diseases, mainly because of the significant homology between mammalian genomes and the many other similarities in physiology and immune components. Moreover, mice are susceptible to an overlapping range of microbes infectious to humans. However, it is important to keep in sight the important differences between humans and other animals and no *in-vivo* model can fully replace investigations in humans or on human materials.

The genetic tools are available for manipulating the mouse [153] (for example, gene knock out technologies and homologous recombination), and the strategies available to control the expression of microbial genes *in-vivo*, make the mouse an excellent experimental model to study the genetic basis of infections. Some of the above-mentioned advantages apply to other mammals, but mice are small and relatively easy to maintain in the laboratory; their short breeding cycle (about 2 months) and their high reproductive capability (5 to 10 offspring per litter and approximately one litter every month) make them suitable for genetic analysis. Mutants mouse lines are becoming available from open sources and new mutations can be introduced using different approaches: irradiation, feeding with chemical mutagens or inserting DNA fragments into the genome using, for example, novel CRISPR/Cas9 type technologies [154]. Further, murine heterologous gene transfer technology is highly advanced in such a way that sophisticated transgenic mice, carrying foreign/heterologous genes of interest (transgenes), have allowed the creation of experimental animal models that further recapitulate aspects of the pathology of human diseases. It is arguably faster to map a mouse disease gene and use its sequence and location to find the position of the ortholog in the human genome, than it is to map the human gene directly.

As a result of recent advances in breeding strategies it is now possible to make congenic mice, which are genetically identical with the exception of being polymorphic in one particular nucleotide, gene or regulatory sequence. In addition, in the mouse, selected genes can be deliberately mutated by swapping the functional copy of the gene for a mutated version in mouse embryonic stem cells (ES cells). This means it is possible to create exact or highly related replicas of the genetic defects that cause diseases in humans [155].

The mouse has proved to be an invaluable model for study many infections, including those caused by *Salmonella*. Many *Salmonella* and host genes that influence the outcome of infection have been identified, including immune genes and some of these have also been shown to influence infection in humans [13]. Some *Salmonella* are highly virulent in the mouse, for example isolates of *S.* Typhimurium and *S.* Enteritidis and murine salmonellosis models have been a cornerstone of studies on pathogenicity and immunity [156]. *S.* Typhimurium can cause an invasive,

systemic disease in mice that resembles aspects of typhoid in humans. *S.* Typhi is relatively avirulent in mice and consequently the *S.* Typhimurium murine infection model has been used as a surrogate for human typhoid. Inbred mice can vary significantly in terms of their susceptibility to *Salmonella* and single gene loci can greatly influence this susceptibility. For example, mice harbouring a protective Nramp1 allele, (for example, 129 lines), are several logs more resistant to *Salmonella* compared to those harbouring equivalent susceptible alleles (for example, C57 Black 6 lines) [157, 158]. Many immune genes also play a key role in protection against *Salmonella* infections; for example, genes encoding TNF-α and Interferon-γ (see [159] for a comprehensive review of this area).

1.7.2.1 The streptomycin-pretreated mouse model for colitis

The most common disease associated with non-typhoidal *Salmonella* in humans is enterocolitis. However in mice, certain serovars including *S.* Typhimurium do not cause gastroenteritis but rather targets the gut-associated lymphatic tissues and cause a systemic typhoid-like infection. To increase the knowledge of the pathogenic mechanisms of intestinal salmonellosis, a streptomycin-pretreated mouse model was established to provide a mouse model for serovar Typhimurium-mediated colitis [160]. Pre-treatment of mice with streptomycin disrupts the natural microbiota, which can limit colonisation by an incoming pathogen [160]. However, mice pre-treated with streptomycin can develop colitis soon after oral infection with *S.* Typhimurium and they present with characteristic symptoms of a human enteric salmonellosis including epithelial ulceration, oedema and infiltration of CD18-positive cells [160]. This pathology is significantly dependent on protein translocation via the *S* Typhimurium SPI-1 secretion system. In addition to colitis, the *S.* Typhimurium can still become systemic in susceptible mice, colonising the liver and spleen.

1.7.3 *In-vivo* model: The zebrafish

Although the mouse is a key model for studying *Salmonella* disease, several aspects of murine biology limit its utility therefore comparative *in-vivo* models are potentially interesting. The zebrafish, *Danio rerio*, has attracted a great deal of

interest as a model of general physiology and infection. Despite the obvious differences in the physiology of fish and humans that could affect the outcome of diseases in the model, the zebrafish offers several features that make it an important complement to mouse models of disease. Among these, the zebrafish provides excellent possibilities for real time *in-vivo* imaging of host-pathogen interactions, given the optical clarity of embryos and larvae. Imaging can be exploited in association with the sophisticated tools for genomic and large-scale mutant analysis available in this species. The small size of the fish, their high reproductive rate (hundreds of offspring per week from a pair), the external development of the embryos along with the low maintenance costs and the establishment of methods to rear embryos under gnotobiotic conditions have contributed to the uniqueness of zebrafish as a model of human diseases. In addition, zebrafish embryos and larvae are highly suitable for screening chemical libraries given their small size and CRISPR/Cas9 technologies have now been applied to this species [161]

A number of reports have already established infection models in the zebrafish using bacterial and viral pathogens including those exploiting *Mycobacteria marinum* [162, 163] and *S. enterica* [164] involving the systemic infection of early embryos. Therefore, the zebrafish model has already been validated to some degree. Zebrafish are also susceptible to parasitic infections and recently, fungal infection models have also been established [164]

The zebrafish, mouse and human share components of both innate and acquired immune systems. Indeed, equivalents of many mammalian immune cells have been identified in zebrafish. Zebrafish embryos possess functional macrophages at day one of development and are capable of sensing and responding to microbial infections. However, innate immune functions can be studied with some degree of separation from adaptive functions in zebrafish embryos, since acquired immune cells develop only later during larval stages and are not fully matured until approximately 4 weeks post fertilisation as depicted in Figure 1.6. However, all major organs are present by 5 days post fertilisation facilitating infection tracking at the tissue level.

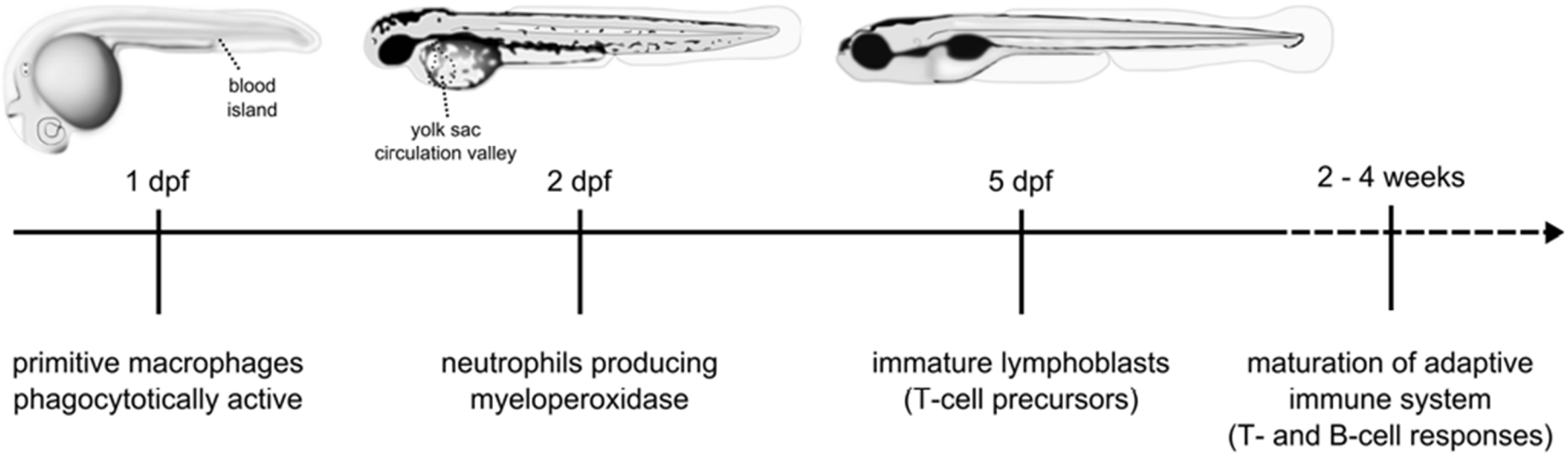


Figure 1.6: Overview of the development of the zebrafish immune system. Taken from [165].

The availability of a zebrafish full genome sequence, efficient tools for forward and reverse genetics and sophisticated mutagenesis and screening strategies on a large scale, and low cost that is not possible in other vertebrate systems also contribute to the usefulness of the model. For forward genetic screens, germ-line mutations were commonly introduced by ethylnitrosourea (ENU) treatment of male zebrafish [166], which yields relatively random point mutations that can be identified by positional cloning and DNA sequencing. Retroviral or transposon-mediated insertion mutagenesis strategies can also be used [167]. Until recently, reverse genetics in zebrafish predominantly relied to an approach known as TILLING (Targeting Induced Local Lesions in Genomes) [168, 169], which compensated for the then lack of conventional knockout technology available for zebrafish. The use of new Zinc-Finger Nuclease (ZFNs) [170] and Transcription Activator-Like Effector Nucleases (TALENs) [170, 171] technologies was introduced as a useful addition to TILLING approaches. Exposure to morpholinos, synthetic oligonucleotides that can be designed to block translation or pre-mRNA splicing, can induce a transient knockdown of gene expression in zebrafish [172].

A more recent approach exploits an old bacterial defence mechanism in which Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR); together with CRISPR- associated (Cas) [173] proteins provide acquired resistance to invading viruses and plasmids. The type II CRISPR/Cas system involves the uptake of foreign DNA fragments into CRISPR loci and subsequent transcription and processing into short CRISPR RNAs (crRNAs), which in turn become a trans-activating crRNA (tracrRNA) and direct sequence-specific silencing of foreign nucleic acid by Cas proteins. Based on previous in vitro studies that had shown that a single synthetic guide RNA (gRNA), consisting of a fusion of crRNA and tracrRNA, can direct Cas9-mediated cleavage of target DNA, a platform exploiting customised RNA-guided Cas9 nucleases has been developed to efficiently induce site-specific modifications *in vivo* in the zebrafish [155] and this approach represent to date the easiest and most efficient way to generate genetic modification of zebrafish. Indeed, the CRISPR/Cas9 and equivalent systems are finding broad utility in mammalian cell genetic manipulation experiments [174]. In terms of genome-based phenotyping, transcriptional responses of zebrafish to infection can be studied by exploiting reverse transcriptase PCR, microarrays and next generation sequencing studies.

However, as this is a relatively new model for infection studies, there is a general lack of immunological reagents for detecting proteins and immune signatures. However, commercial ventures in antibody production for zebrafish such as ANASPEC are increasing. In addition, sometimes, antibodies to the mammalian orthologs of some zebrafish proteins show cross-reactivity. The readily available zebrafish mutant library at the WTSI, which now covers more than a third of the total protein coding genes of the genome, is potentially available to explore pathogen-host interaction in specific infection models. Potential infection challenges include *M. marinum* and *S.* Typhimurium, which can efficiently infect zebrafish and can be a resource to characterise novel infection susceptibility loci.

1.7.4 The mouse genetic screening

The WTSI is conducting a large phenotypic screen on novel mice harbouring defined and conditional ready mutations in different genes. The mice are of the C57/B6 lineage. To generate these mice, heterozygous ES cells harbouring specific mutations are selected for microinjection from a library of over 15,000 mutant stem cells [175]. At the time of writing this thesis, over 1,250 mice lines have been screened for a range of phenotypes, including plasma chemistry and infection susceptibility. Particular mutant lines are selected on the basis of a number of criteria including novelty, whether they are a hit in previous GWAS studies or exome sequencing, or through recommendation. Currently for infection susceptibility, mice are challenged independently with *S.* Typhimurium M525, influenza virus X31, the worm *Trichurius muris* and DSS (Dextran Sodium Sulphate). The responses of pathogen challenged mutant mice are compared to similarly infected wild type mice to assess the influence of the murine mutation on phenotype. The combined data can be presented as a heat map, which summarises phenotypic differences found in any mutant mouse line (http://www.mousephenotype.com).

1.7.5 The zebrafish genetic screening

The Zebrafish Mutation Project (ZMP) at the WTSI has an ultimate goal to create a mutant allele in every protein-coding gene in the zebrafish genome, using a combination of whole exome enrichment and Illumina next generation sequencing.

To date, over 26000 alleles have been generated. To exploit the mutant library generated, a high-throughput, systematic phenotypic analysis has been developed to assess the phenotype associated with any given mutation with a particular attention to nonsense and essential splice mutations.

Secondly, correlations between the predicted disruptive mutations and phenotype are established by crossing heterozygous adult fish then examining the embryos for morphological and behavioural phenotypes at 5 days post fertilisation (dpf). Phenotypes can be linked to genotyped my successive rounds of sequencing and genetic crossing. After a particular mutation has been described further phenotyping or genome re-engineering can be attempted. Transcript counting or RNA-seq can also be performed as part of the primary or secondary phenotyping (http://sanger.ac.uk/resources/zebrafish/zmp/).

1.8 The use of whole genome sequencing for studying bacterial genomes and phylogeny

During the last century combined studies began to highlight the huge genetic diversity within the bacterial world and leading scientists started to develop phylogenetic schemes in an attempt to explain how life on Earth may have developed. An application of phylogenetic analysis is to predict ancestral structures that can help to understand the evolutionary path of organisms. The first phylogenetic trees of prokaryotes were largely based on morphological, physiological and biochemical analysis. The prokaryote-eukaryote dichotomy was already well established but did very little to clarify phylogenetic relationships. In 1970, Carl Woese used a molecular approach to phylogenetics, arguably transforming our understanding of evolution in the microbial world by introducing the Archaeal domain [176, 177].

With the advancements in DNA sequencing technologies and the associated increase in nucleic acid sequence information, the application of phylogenetic analysis has rapidly expanded. It has now been applied broadly to multiple bacterial species as well as serovars or clades within the same species [178-181]. Comparative genomics

and phylogenomics have been exploited to trace the emergence of new drug-resistant bacterial clades [181], tracked the global spread of infectious pathogens [181, 182] as well as identify the pandemic source of some infectious threats [183-185]. The use of phylogeny also enables a clearer understanding of how bacterial genomes evolve and adapt to novel selective pressures. For example, studies have shown that patterns of antibiotic usage influences bacteria genome evolution and the patterns of recombinant signatures maintained within the genome [186].

Modern phylogenetic analysis relies on the availability of high quality genome sequence information. In order to meet such needs, several new nucleic acid sequencing technologies have been developed over the past few years. The Sanger sequencing method was the first widely used sequencing technology to be exploited for bacterial genome sequencing [187, 188]. The first complete bacterial genome to be fully sequenced by this approach was that of *Haemophilus influenze* but many more followed in the following decade [189]. The Sanger sequencing system exploits the addition of terminating dideoxyribonucleotides (ddNTPs) by DNA polymerase, preventing the incorporation of further nucleotides.

Although the Sanger method found wide utility, it was not particularly high throughput and was a relatively expensive approach. Next generation sequencing platforms were invented that exploit the immobilisation of DNA samples onto solid supports, incorporate automated cyclic sequencing reactions mediated by fluidics devices and exploit sensitive detection of molecular events by imaging. These revolutionary technologies are capable of producing an enormous amount of sequence data in a relatively short period of time while keeping cost relatively low [190]. They can also now generate relatively long DNA sequences facilitating genome assemblies [191].

Two new generation sequencing platforms based on the Illumina/ Solexa and Pacific Biosciences technology, were utilised in the studies described in this thesis. The Illumina sequencing platform exploits sequencing by synthesis method in which modified dNTPs containing a fluorescently labeled reversible terminator blocks further polymerisation so that only a single base can be added by a polymerase enzyme to the DNA copy strand. The terminator is imaged then cleaved off to allow

incorporation of the next base. The sequencing reaction is conducted simultaneously on a large number of different templates spread onto a solid surface, forming clusters. The natural competition between all 4 modified dNTPs present during each sequencing cycle minimises incorporation bias. Illumina sequencing platforms provide multiple applications from whole genome sequencing, SNP (single nucleotide polymorphism) detection to transcriptomics and metagenomics analysis. The HiSeq, Nextseq and GAIIx platforms are suited for studying larger genomes (animal or plant) while the MiSeq platform is ideal for small genomes or targeted regions within a genome. The HiSeq X Ten platform (2014 release) is limited to sequencing only whole genome human samples. Illumina platform limitations include inadequacies in analysing low diversity samples and they have relatively short reads compare to other platforms, although this is improving with longer reads now achievable (https://www.illumina.com/).

The Pacific Bioscience sequencing system, also known as PacBio RS/RS II (latest release, 2014) exploits a Single Molecule Real Time (SMRT) method in which an optical waveguide (the zero-mode wave guide) is attached to the DNA polymerase, generating an illuminated observation volume, small enough to observe the addition of only one single nucleotide. Each nucleotide is attached to a different fluorescent dye and when incorporated through the DNA polymerase, the fluorescent tag of the nucleotide is cleaved off and is no longer observable within the optical waveguide area. The fluorescent signal of the nucleotide is detected and the corresponding base call is made according to the fluorescence of the specific dye. The major benefits of using PacBio sequencing technologies have been attributed to the production of reads significantly longer than other sequencing platforms making it ideal for sequencing small genomes (such as bacteria or viruses) and assembling larger genomes. Also, the system can facilitate the sequencing of regions of high G/C content and can identify some modified bases (methylation, hydroxymethylation) without necessitating the need for chemical conversion during library preparation (http://www.pacificbiosciences.com/).

1.9 The focus of this thesis, *S*. Weltevreden

Salmonella enterica serovar Weltevreden has recently attracted a great deal of interest due to increasing reports of its isolation by reference laboratories and other microbiological centres world-wide, particularly in Asia. *S.* Weltevreden has been associated with potential marine sources and it plays a significant role in food poisoning. Indeed, a global *Salmonella* survey conducted by WHO revealed that this organism is the most common cause of non-typhoidal salmonellosis in the South East Asian Region (SEAR) and Western Pacific region [192]. It is frequently isolated from seafood, meat, poultry products and water. Prevalence of *S.* Weltevreden was detected in domestic animals like pigs, chicken and ducks in Vietnam and it is also the most common serovar isolated from humans in Thailand and Malaysia. According to the *Salmonella* food poisoning database during 1989-99, *S.* Weltevreden was the second common pathogen encountered, next to *S.* Enteritidis [193].

Despite the emergence of *S.* Weltervreden as a significant health problem relatively little has been reported about the genotypic or phenotypic properties of this understudied serovar. Reports have been emerging providing the first genome sequence data of individual isolates (see chapter 3) but to date these have not been placed in a phylogenetic or evolutionary context.

1.10 Aims and objectives of thesis

In this study, whole genome sequencing technologies linked to phylogenetics and comparative genomics were used to define the genetic diversity within a large collection of *S.* Weltevreden isolates collected worldwide from diverse sources, with a focus in Vietnam where such infections are common. This focus on isolates from South East Asia and Western Pacific region provide an opportunity to explore the relationships between *S.* Weltevreden predominance in this region. Phenotypic characterisation was performed on selected isolates, with an aim to dissect aspects of host-pathogen interaction during infection, providing a foundation to compare *S.* Weltevreden with more commons enterics. Thus, the aims of this thesis were to:-

- Define the phylogenetic structure of the *S.* Weltevreden serovar, define the core and accessory genomes and provide a high quality reference genome.

- To assess the pathogenic and metabolic potential of *S.* Weltevreden using simple laboratory assays, including both *in-vitro* (cellular) and *in-vivo* (mouse, zebrafish models) virulence assays.

2 Materials and methods

2.1 Phylogenetic analysis of *S*. Weltevreden

2.1.1 Illumina sequencing

All DNA samples were processed and sequenced by the core sequencing facilities at the WTSI. Multiplex libraries were generated using DNA insert of ~200 to 300bp with each isolate uniquely tagged. Samples were sequenced using the Illumina Hiseq platform (Illumina, Inc., San Diego, California USA) to produce ~100bp reads. The first stage of the library preparation involved DNA fragmentation by focused ultrasonication using a Covaris E-series ultrasonicator (Covaris, Inc., Woburn Massachussett, USA). This was followed by DNA purification using the magnetic bead-based technology solid Phase Reversible Immobilisation (SPRI) from Agencourt Bioscience (Agencourt Bioscience Corporation, Beverly, Massachussett, USA). After this stage, small fragments were removed and the remaining DNA consisting of a mixture of blunt end fragments, were repaired. A single "A" nucleotide moiety was added to the 3' ends of the fragments followed by successive adaptor ligation. This step, called A-Tailing, deters concatemerisation of templates and increased the efficiency of adaptor ligation. Specific adaptors were ligated to the 3' and 5' ends of the DNA templates.

The DNA molecules that were correctly attached to the adaptors were amplified using the DNA polymerase Kapa HiFi enzyme (Kapa Biosystems, Woburn Massachusetts, USA) and primers that targeted the unique library index tag. Amplification completed the construction of the adaptor ends to produce a fully double stranded template. This PCR-amplified library was then denatured using sodium hydroxide in hybridisation buffer at a concentration of 3.5pM in order to create single stranded DNA, which was loaded onto a single lane of the flowcell on the Illumina Hiseq platform. Protocols for cluster formation, primers hybridisation

and paired-end sequencing reaction were performed according to the manufacturer's recommendations.

2.1.2 Sequence assembly from Illumina reads

The Illumina-generated sequences were assembled using a pipeline (https://github.com/sanger-pathogens/vr-codebase) developed at the WTSI. For each genome, the *de novo* short-read assembler Velvet [194] (version 1.2.09) was used to generate multiple assemblies by varying the k-mer size between 66% and 90% of the read length using Velvet Optimiser (https://github.com/tseemann/VelvetOptimiser). From these assemblies, the assembly with the highest N50 was chosen. Contigs were excluded from the assembly if they were shorter than the target fragment size (400 bases).

An assembly improvement step (https://github.com/sanger-pathogens/assembly_improvement) was then run. The raw reads were mapped to the assembly and reads which mapped in perfect pairs to the same contig, were excluded (since these have already been successfully used). The remaining unmapped and mapped reads were used in an improvement step to try and reduce the fragmentation of the assembly. A scaffold assembly of the contigs was built by iteratively running SSPACE [195] (version 2.0) beginning with the contigs which were predicted to map next to each other. The reads were then mapped again to the scaffold assembly and perfect pairs were excluded. Next, gaps identified as one or more N's, were targeted for closure by running 120 iterations of GapFiller [196] (version 1.11), using a decreasing read evidence threshold. Finally, the reads were aligned back to the improved assembly using SMALT (https://www.sanger.ac.uk/resources/software/smalt/) and a set of statistics was produced for assessing the quality of the assembly. All the assemblies produced were created in a standardised manner and required no input from the user so all the results are reproducible. The median number of contigs for the sample set was 68.

2.1.3 Pacific Biosciences (PacBio) assembly

The PacBio raw read data for each sample was manually assembled by Dr. Martin Hunt (WTSI Pathogen Informatics team) using the PacBio SMRT analysis pipeline (https://github.com/PacificBiosciences/SMRT-Analysis) (version 2.2) utilising the HGAP assembler [197]. The raw unfinished assemblies all produced a single uncircularised chromosome plus some other small contigs, some of which were plasmids or unresolved assembly variants. If the ends of a contig overlapped, they were identified as candidates for circularisation using a protocol recommended by PacBio (https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/Circularizing-and-trimming). Figures 2.1.a and 2.1.b illustrate this process. A virtual break was manually introduced into the chromosome sequence at the *thrA* gene, to match the starting point of other published *S. enterica* references. Plasmids were also artificially broken at the replication gene. The sequences were then circularised using the genome assembler, Minimus [198] (version 2 part of AMOS version 3.1), which removed the overlapping sequence. Quiver was then used by the circularised sequence and the raw reads to correct errors in the circularised region. As high quality short read data from Illumina were available, ICORN2 (Otto et al. 2010) (version 0.97) was used to correct minor errors in the assembly, providing a very high quality reference sequence, as assessed by REAPR [199] assembly was subsequently annotated with Prokka [200].

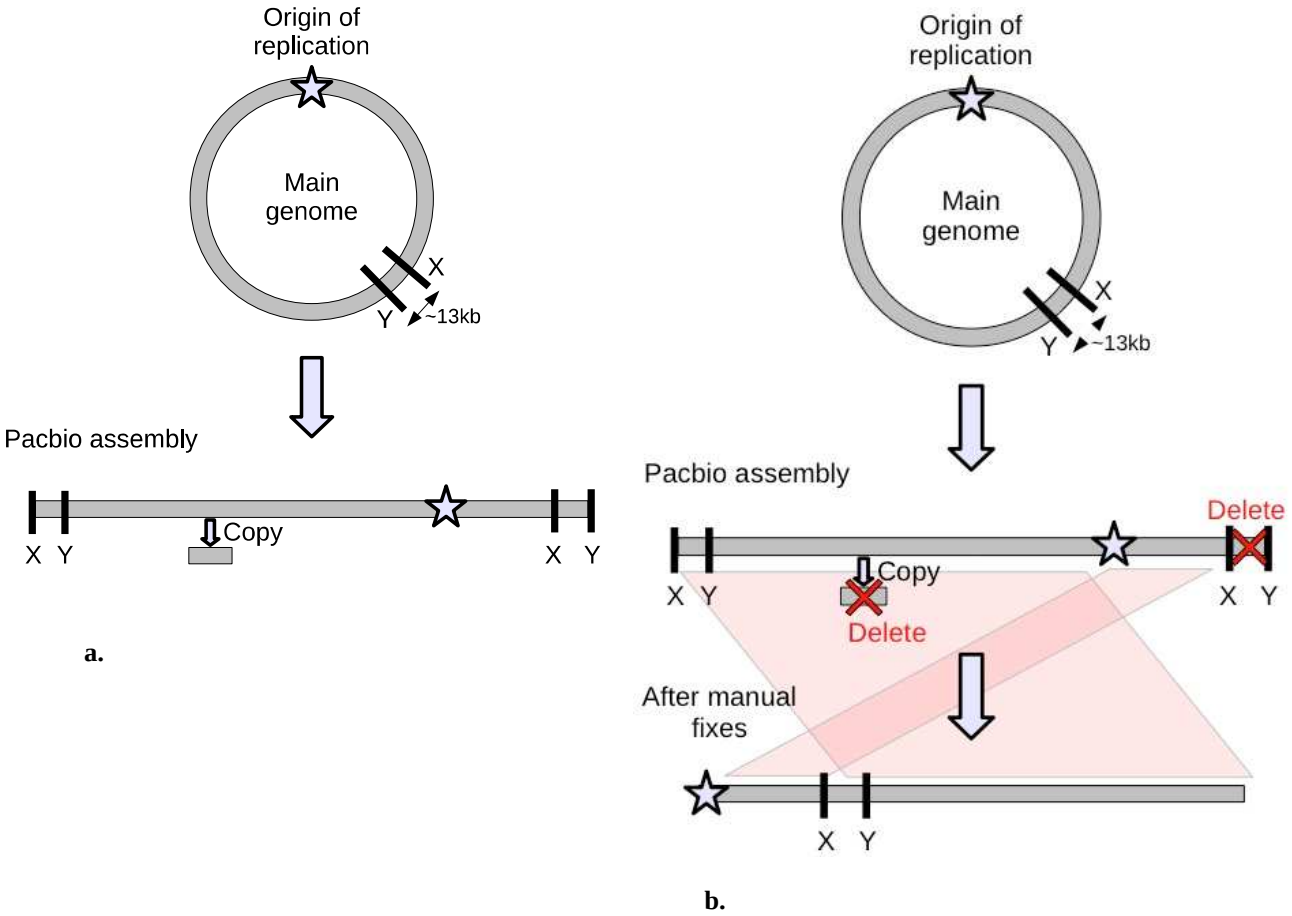


Figure 2.1: a. graphic representation of S. Weltevreden genome assembly before manual fixing. b. Graphic representation of S. Weltevreden assembly before and after manual fixing.

2.1.4 Annotation

Each *de novo* assembly was annotated using Prokka in an automated pipeline (https://github.com/sanger-pathogens/Bio-AutomatedAnnotation). Coding regions were first predicted using Prodigal [201] and tRNA/tmRNA genes using ARAGORN [202]. These were then annotated using a number of databases involving a combination of protein blast [203] and HMMER. To provide community specific gene names and annotation, a *Salmonella* database of amino acids was generated from all of the annotated *Salmonella* genomes in RefSeq (http://www.ncbi.nlm.nih.gov/refseq/). Prokka is bundled with prebuilt databases from UniprotKB (SwissProt), clusters, CDD, TIGRFAMs, PFAM (A) and RFAM which provide more general high quality annotation at the family level.

As *S.* Weltevreden 10259 was sequenced on both PacBio and Illumina we could perform further analysis using a short read assembly versus a longer read finished assembly. The Illumina assembly contained 5034 coding regions and the PacBio assembly contained 5110, giving an under prediction of 1.4%. For the whole dataset the median number of genes predicted from the short reads assemblies was 4902.

2.1.5 MLST from *de novo* assemblies

The MLST results were verified from the *de novo* assemblies using MLST check (https://github.com/sanger-pathogens/mlst_check/) (version 2.0.1510612). All of the assemblies were blasted against the *S. enterica* MLST database (Achtman et al. 2012) and were verified as being ST 365. These data were also checked for the presence of novel alleles (there should only be 1 copy), a process which can also highlight contamination from a closely related serovar, but no obvious contamination was detected.

2.1.6 Checking for *S.* Weltevrden in sequencing reads

A Kraken [204] database (version 0.10.6) was created containing the reference genomes for Homo sapiens (GRCh38), Mus Musculus (GRCm10), and all

archea/viruses/plasmids/bacteria in RefSeq (http://www.ncbi.nlm.nih.gov/refseq/). For every sample, each read was categorised to a taxonomic identifier, with the results collated into a single report for manual inspection. Any contaminated samples were excluded from further analysis. This check highlighted one isolate, which was contaminated with *E. coli* DNA and another which was a mixture of *Salmonella* serovars. No host contamination was identified.

2.1.7 Detecting regions likely to be erroneous with short read sequencing

The median fragment size was 400 bases for the isolates sequenced on Illumina. Repeats larger than this size cannot be reliably resolved, thus any variants which fall into these regions cannot be trusted. This is a fundamental limitation of short read sequencing technologies and cannot be resolved used bioninformatics. Consequently, the reference genome was blasted against itself [203] (version 2.2.31) and the coordinates of matches (query and reference), which were over 400 bases in length and had greater than 99% identity were noted. All bases falling within these regions were then replaced with 'N' in the multi-FASTA alignment file.

2.1.8 Recombination mapping

The filtered multi-FASTA alignment was then checked for recombination using Gubbins [205] (version 1.3.4). Five iterations of Gubbins were run and in each iteration a phylogenetic tree was constructed with RAxML [206] (version 7.8.6) with the GAMMA GTR model, internal ancestral sequences were inferred using FastML [207] (version 3.1). Recombinant sequences were detected and a multi-FASTA alignment with the recombinant regions was masked out. This data was then used as the input to the next iteration. RAxML with 100 bootstraps was then run over the final multi-FASTA alignment to provide a high quality phylogenetic tree in newick format.

2.1.9 General mapping

All of the Illumina sequencing short read data was aligned to the complete reference genome generated from *S.* Weltevreden 10259. The reads, in FASTQ format, were first split into chunks of one million reads. Each chunk of reads was individually aligned using SMALT (https://www.sanger.ac.uk/resources/software/smalt/) (version 0.7.4), a hashing based read aligner. The aligned reads are then merged together using samtools [208] (version 0.1.19), coordinate sorted, and outputted as a BAM file. Optical duplicates were identified using Picard (http://broadinstitute.github.io/picard/) (version 1.9.2). Statistics about each mapping were generated using BamCheck [208] (version 0.1.19, but since renamed as 'samtools stats') including read coverage of the reference genome, reads aligned, perfect pairs, unmapped reads, actual insert size, etc. and these results were evaluated manually to identify poor quality sequencing data.

2.1.10 SNP calling

SNPs were called on each set of aligned reads using mpileup with the parameters 'samtools mpileup -d 1000 -DSugBf ref bam'. The raw SNPs were then passed into BCFtools and were filtered into a higher quality set. A virtual pseudo-genome was then constructed by substituting the base call at each site (variant and non-variant) into the reference genome. For a SNP to be called the depth had to be greater than 4 reads, and be present on both strands, with at least 75% of reads containing the SNP at that position. The mapping quality had to be greater than 30 (less than 1 in 1000 probability that the mapping was incorrect). If a SNP failed to meet these criteria it is substituted with an 'N'. Insertions with respect to the reference genome were ignored. Deletions with respect to the reference genome were filled up with 'N' characters in the pseudo-genome in order to keep it aligned and at the same length relative to the reference genome. Heterozygous sites were turned into homozygous alleles by selecting the first allele in the BCF file. However, if the first allele was an insertion or deletion (indel), the second allele in the BCF file was taken. If the second allele was also an indel, a single 'N' character was used. All of the pseudo-

genomes were then merged into a single multi-FASTA alignment file, including the reference sequence.

2.1.11 Clusters and defining SNPs

The population structure of the phylogenetic tree was validated using a Bayesian statistical approach. Hierarchical BAPS [209] (version 6.0 of BAPS) was used to perform a hierarchical clustering of the multi-FASTA alignment (after recombination's had been removed) to reveal a nested genetic population structure. Two distinct *S.* Weltevreden clusters were identified by BAPS. SNPs, which uniquely (in 100% of isolates in a cluster), defined each of the clusters were extracted using BioPericles (https://github.com/sanger-pathogens/BioPericles) (version 0.1.0). Exploiting the multi-FASTA alignment with recombination removed, a consensus sequence was generated for each cluster and any bases which varied or contained missing data were replaced by 'N'. The consensus sequences were merged into a single multi-FASTA alignment file and SNP locations were identified using SNP sites (https://github.com/sanger-pathogens/snp_sites) (version 2.0.1). Each SNP was then annotated using the reference annotation (10259) GFF3 file. An annotated VCF file was produced with VEP syntax [210] listing the type of change (intergenic/ synonymous/ nonsynonymous), the amino acid (before and after) and the amino acid position in the gene, along with the coordinates of each SNP relative to the reference genome, the reference base, the allele base and the presence and absence of the variant in each cluster. These cluster defining SNPs were then further annotated with the functional annotation of the gene they occurred in.

2.1.12 Predicting antibiotic resistance

Antibiotic resistance was predicted from each sample's raw sequencing reads using ARIBA [211] (version 0.4.1), which performs antibiotic resistance identification by assembly and alignment. A manually curated input database of known resistance genes in FASTA format was used as input along with the paired end sequencing reads in FASTQ format. The resistance sequences were first clustered using CD-hit [212] (version 4.6). The raw reads were then aligned to a representative sequence for each resistance cluster. Reads which mapped and their complimentary strand

equivalents (even if unmapped) were extracted. A local assembly was performed on the reads for each cluster using [213] (version 3.5), where the resistance genes for the cluster were used as ‘untrusted contigs’. This generates a candidate gene along with sequence on either side if the gene is present in the reads. Algorithms which only use alignment suffer from a coverage drop off at either end of the gene, making identification less reliable, such as was found with SRST2 [214]. MUMmer [215] (version 3.23) was then used to identify differences between the assembled contig and the known resistance gene and the results were reported along with any variation found and quality flags. These were manually inspected and samples with 100% matches to resistance genes and with a complete open reading frame were flagged as being potentially candidates for visual inspection.

2.1.13 Pan genome analysis

A pan genome was constructed using Roary [216] (version 3.2.5) from the annotated assemblies of the sample set with a percentage protein identity of 95%. This first step identified both the candidate core genes, conserved across all isolates and the accessory genes, which vary across isolates. The protein sequences were first extracted and iteratively pre-clustered with cd-hit (version 4.6) down to 98% identity. An all against all blast (version 2.2.31) was performed on the remaining unclustered sequences and a single representative sequence from each cd-hit cluster was selected. The data were used by MCL [217] (version 11-294) to cluster the sequences. The preclusters and the MCL clusters were merged and paralogs were split by inspecting the conserved gene neighbourhood [218] around each sequence (5 genes on either side). Each sequence for each cluster was independently aligned using PRANK (Löytynoja 2014) (version 0.140603) and combined to form a multi-FASTA alignment of the core genes.

2.2 Phenotypic characterisation of *S.* Weltevreden

2.2.1 Bacterial strains culture conditions

All bacteria were routinely grown on Luria-Bertani agar (LB agar) plates and broth (LB medium) at 37°C. For in vitro assays, genetically transformed isolates harbouring the plasmid pSsaG (pSsaG directs the expression of green fluorescent protein (GFP) from the *ssaG* promoter) [219, 220] were used. LB agar and broth were supplemented with ampicillin at a concentration of 100 μg/ml. For zebrafish infection challenges, bacteria harbouring the plasmid pRTZ3, which drives the constitutive expression of the red fluorescent protein dsRed, were used. For these experiments, LB agar and broth were supplemented with tetracycline at a concentration of 30 μg/ml.

S. Weltevreden isolates C2346, 10259, 98 11262 and 99 3134 were used throughout for phenotyping. *S.* Typhimurium, used as a control, SL1344 was provided by Dr. Derek Pickard from the Wellcome Trust Sanger Institute. *S.* Weltevreden C2346 and 10259 were obtained from OUCRU Ho Chi Minh City Vietnam. *S.*Weltevreden 98 11262 and 99 3134 were supplied by the Centre National de Référence *E. coli*/*Shigella*/*Salmonella*, Unité de Recherche et d'Expertise des Bactéries, Pasteur Institute, France.

2.2.2 Serological identification

The identification and confirmation of the serotype of each *S.* Weltevreden isolate was performed by a standard agglutination test using O, H or Vi antisera. *S.* Weltevreden is classified as O3, O10 or O15 positive; R and Z6 positive and Vi negative based on the Kauffman-White scheme [221]. Anti-*Salmonella* O3 mouse antibody, anti-*Salmonella* O4 mouse antibody (negative control), anti-*Salmonella* O10mouse antibody, anti-*Salmonella* O15mouse antibody, anti-*Salmonella* Hr mouse antibody, anti-*Salmonella* Hz6 mouse antibody and anti-*Salmonella* Vi mouse antibody were obtained from Sifin and rabbit anti-*Salmonella* O1,3,19 from Statens Serum Institute (Copenhagen, Denmark) were used for agglutination tests. A single

colony was mixed independently with each of the 7 antibodies and visible clumping (agglutination) was observed within 2 minutes for a positive reaction.– was used to designate no agglutination through to ++++ for a strong agglutination.

2.2.3 Bacterial growth assessment

An isolated colony of *S*. Typhimurium SL1344 or *S*. Weltevreden C2346, 10259, 98_11262 and 99_3134 was grown overnight in 10 ml of LB broth at 37°C. The following day, 50 ml of a secondary culture was started from the overnight at an OD_{600} of 0.05. Experiments were performed at 37°C and 28°C and the OD_{600} was measured each hour for the first 2 hours then each 30 minutes for the rest of the assay. At each time point 1 ml of culture was used in serial dilution for subsequent plating on LB plates to assess bacterial growth.

2.2.4 Microarray assay (Biolog)

Phenotype microarrays (PM) to assess the metabolism of individual carbon sources (PM 1 to 2), nitrogen sources (PM 3), phosphorus and sulphur sources (PM 4), biosynthetic pathway substrates (PM 5), osmotic/ionic response (PM 9) and pH response (PM 10) were performed according to the manufacturer's instructions. (Biolog Inc. Hayward, California, USA). The bacteria were grown up to OD_{600} 0.667 and the cell suspensions were made up to a transmittance of 42%. For *S*. Typhimurium SL1344 the cell suspension was supplemented with histidine and the carbon source used was succinate. For *S*. Weltevreden C2346 the cell suspension was supplemented with adenosine and the carbon source used was succinate while for the isolates 10259, 98_11262 and 99_3134 the carbon source used was pyruvate with no added supplement. PM micro titter plates were incubated at 37°C for 48 hours in the Omnilog (Biolog Inc.) and each well was monitored for redox indictor change representing kinetic respiration. Tests were performed in duplicate and the kinetic data analysed using Omnilog PM software (Biolog Inc.). Data was exported from the Biolog File Manager, and further analysis was conducted in R.

2.2.5 Gentamicin killing assays using Hep2 cells

To facilitate the analysis of invasion assays, *S*. Typhimurium SL1344 and *S*. Weltevreden C2346, 10259, 98_11262 and 99_3134 were transformed with the plasmid pSsaG that directs the expression of GFP from the *ssaG* promoter [219, 220]. Hep 2 cells were cultured in Glasgow's minimal essential medium (GMEM, Sigma) supplemented with 2 mM L-Glutamate and 10% (volume/volume) heat-inactivated fetal bovine serum (FBS). Cells were seeded into 24-well plates (10^5 cells per well) and cultured overnight. *Salmonella* were initially cultured at 37°C with shaking (250 rpm) in 5 ml LB broth for 4.5 h. An aliquot was then diluted 1:50 in L broth and grown at 37°C overnight as a static culture to optimise *Salmonella* pathogenicity island 1 (SPI1) gene expression. For infection, the bacterial cultures were re-suspended in fresh GMEM supplemented with 2 mM L-Glutamate and 10% (volume/volume) heat-inactivated FBS, in order to obtain a multiplicity of infection (MOI) of 50. The MOI was confirmed by plating 10 μl spots of 10-fold serial dilutions of the bacterial solution onto agar plates. After 30 min of incubation (to allow *Salmonella* invasion), cells were washed with phosphate-buffered saline before adding GMEM supplemented with gentamicin (50 μg/ml). Cells were incubated for the appropriate length of time and then washed and lysed with 0.1% Triton X-100. Dilutions of the cell lysates were plated onto agar plates to determine the number of intracellular bacteria. Alternatively, cells were washed and fixed onto 13mm coverslips with 4% formaldehyde then stored in PBS for confocal or electron microscopy. This protocol was adapted from **[222]**

2.2.6 Confocal microscopy

Salmonella infected cells were washed twice with the wash buffer from the Cytotoxicity 3 kit after fixation and permeabilised with the permeability buffer from the same kit for 10 minutes. The cells were then blocked with the block buffer for 20 min at room temperature and stained with goat anti-*Salmonella* CSA-1 antibody followed by tagged secondary antibody. Glass coverslips were mounted onto a microscopic slide along with ProLong Gold antifade reagent DAPI (Invitrogen). The preparations were observed with an LSM510 META confocal microscope (Zeiss).

2.2.7 Scanning Electron Microscopy

After infection with *S.* Typhimurium SL1344 or *S.* Weltevreden C2346, 10259, 98_11262 or 99_3134 for 30 minutes (MOI=100), cells were fixed directly on glass coverslips with 2.5% glutaraldehyde and 4% paraformaldehyde in 0.01 M PBS at 4°C for 1 hour, rinsed thoroughly in 0.1 M sodium cacodylate buffer 3 times, and fixed again in 1% buffered osmium tetroxide for 2 hours at room temperature. To improve conductivity, using the OTOTO protocol devised by Malick and Wilson [223], the samples were then impregnated with 1% aqueous thiocarbohydrazide and osmium tetroxide layers, with the steps separated by sodium cacodylate washes. The coverslip preparations were dehydrated 3 times using an ethanol series (30, 50, 70, 90, and 100% ethanol, 20 minutes each) before they were critical point dried in a Leica CPD300 and mounted on aluminium stubs with conducting silver. Before a specimen had completely set, the coverslip was broken by applying pressure with a sharp point to the centre, which caused radial fragmentation of the glass, in order to obtain better conductivity between the stub and the cells. The coverslip was then sputter coated with a 2-nm gold layer in a Leica ACE600 and examined with a Hitachi SU-8030 SEM.

2.2.8 Transmission Electron Microscopy

Cells were infected as described for the scanning electron microscopy and fixed on ice in their culture wells with a mixture of 2.5% glutaraldehyde and 4% formaldehyde in PBS for 1 hour. The cells were rinsed 3 times with 0.1 M sodium cacodylate buffer (pH 7.42), carefully removed from the plate with a Teflon scraper, and centrifuged at 10,000 rpm for 5 minutes. The pellet was post fixed in buffered 1% osmium tetroxide at room temperature for 1 hour, followed by 1% buffered tannic acid for 30 minutes and then a 1% aqueous sodium sulphate rinse for 10 minutes. The sample was then dehydrated using an ethanol-propylene oxide series (with 2% uranyl acetate added at the 30% step) and embedded in Epon- araldite for 24 hour at 60°C. Ultrathin sections (60 nm) were cut with a Leica EMUC6 ultramicrotome, contrasted with uranyl acetate and lead citrate, and viewed with an FEI 120-kV Spirit Biotwin TEM. Images were obtained with a Tietz F415 digital

TemCam.

2.2.9 Murine intravenous challenge

Three groups of five C57BL/6 mice were challenged intravenously with 2 x 10^3 colony forming units of respectively *S.* Typhimurium SL1344, *S.* Weltevreden C2346, and 10259 as described in [224] . The mice were followed for 4 days checking for survival. They were all subsequently culled at day 4, or earlier if they were critically moribund.

2.2.10 Colitis infection challenges

Six groups of five C57BL/6 mice each were pre-treated with 10 mg of streptomycin (200 µl of a stock solution of 50 mg/ml of streptomycin) 24 hours before challenge. The first group (naïve) was injected with PBS; the second, the third and the fourth groups were infected with approximately 5,5x 10^5 CFUs of respectively *S.* Typhimurium SL1344, *S.* Weltevreden C2346, *S.* Weltevreden 10259, *S.* Weltevreden 98_11262 or *S.* Weltevreden 99_3134. The mice were sacrificed 4 days post challenge and caecum and liver were removed from all mice for further analysis. This protocol was adapted from **[160].** The liver was plated on LB plate for CFU counts in order to check for systemic disease. Part of the caecum was used for histology to look for inflammation and the remaining part was plated on LB agar in order to check for bacterial colonisation of the colon.

2.2.11 Histology

Mice caecum segments were fixed in 4% paraformaldehyde; 5 µm-thick paraffin sections were stained in haematoxylin and eosin according to standard protocols. Stained section were analysed under microscopy to look for sign of intestinal inflammation.

2.2.12 The zebrafish challenge model

Groups of 30 to 50 embryos of AB wildtype zebrafish were challenged intravenously (blood island) with 250 to 300 colony-forming units of *S.* Typhimurium SL1344, *S.* Weltevreden C2346 or mock injected with PBS and phenol red solution. The embryos were followed for 70 hours post infection removing any dead bacteria at each time points. The results were subsequently reported in a survival graph. The same set up was used in a secondary experiment were 10 embryos per time point were first homogenised using a stomacher then plated on LB-tetracycline plate for CFU counts to check for bacterial replication within the host. For experiments involving zebrafish harbouring mutations in Irf8_st95 [225], groups of 30 to 50 embryos from a cross of heterozygous Irf8_st95 and wild type zebrafish parents were challenged intravenously (blood island) with 250 to 300 colony forming units of *S.* Typhimurium SL1344, *S.* Weltevreden C2346 or mock injected with PBS and phenol red solution. 70 hours post infection, the surviving embryos were genotyped and the results were reported and compared to the expected ratio of a quarter wild type, a quarter homozygous and a half heterozygous.

3 Phylogenetic diversity within *S.* Weltevreden

3.1 Introduction

The recent advancements in genome sequencing technologies have enabled a better understanding of the basis of bacterial evolution and adaptation to selective pressures. Such sophisticated technologies have facilitated outbreak tracking and the identification of transmission events in clinical settings as well as in the community. These techniques have been successfully used to address the evolution of various *Salmonella* serovars such as *S.* Typhimurium [226], *S.* Enteritidis [227] and *S.* Typhi [133]. Despite a global effort to study and tackle these pathogens, salmonellosis still represent an important health concern globally as new strains emerge over time. For example, *S.* Weltevreden has emerged as a significant foodborne pathogen particularly in South-East Asian countries and the Pacific region. *S.* Weltevreden has increasingly been reported to be associated with different sources including vegetables, poultry, meat, animal feed and seafood. Indeed, in a study conducted on over 12,000 *Salmonella* isolates, *S.* Weltevreden was the most frequently isolated serovar from seafood in Vietnam and was amongst the highest *Salmonella* contaminant in fish and seafood samples in the world [228, 229]. Various reports of food poisoning due to *S.* Weltevreden have come from India [230], Reunion Island [231], Thailand, Vietnam [232], Fiji [233] and more recently from Norway, Denmark and Finland [234].

In recent years, multiple cases and outbreaks of *S.* Weltevreden have been reported to be associated with different disease outcomes. The clinical outcomes in patients range from asymptomatic carriage, moderate to severe diarrhoea, invasiveness in immunocompromised individuals [235], ulcerative skin lesions [236], through to rare cases of fatality [237]. Antibacterial resistance is not currently commonly reported for *S.* Weltevreden isolates and current therapy includes the use of fluoroquinolones.

Some of the earliest attempts to genetically characterise *S*. Weltevreden involved the generation of draft genomes and preliminary comparative genome analysis. Draft genomes have been generated for the scallop-associated isolate SL484, the Scandinavian outbreak associated isolate *S*. Weltevreden 2007-60-3289-1 from alfalfa sprouts [238] and a multidrug resistant isolate *S*. Weltevreden "9" isolated from seafood [239]. Preliminary comparative genomic studies performed against representative isolates of *S*. Dublin, *S*. Newport, *Salmonella* Cholereasius, *S*. Enteritidis, *Salmonella* Gallinarium, *Salmonella* Heidelberg, *Salmonella* Agona, *S*. Paratyphi (A, B and C), *Salmonella* Schwarzegrund, and *S*. Typhi [238] were used to generate a phylogenetic tree showing their comparative relationships to *S*. Weltevreden (Figure 3.1).

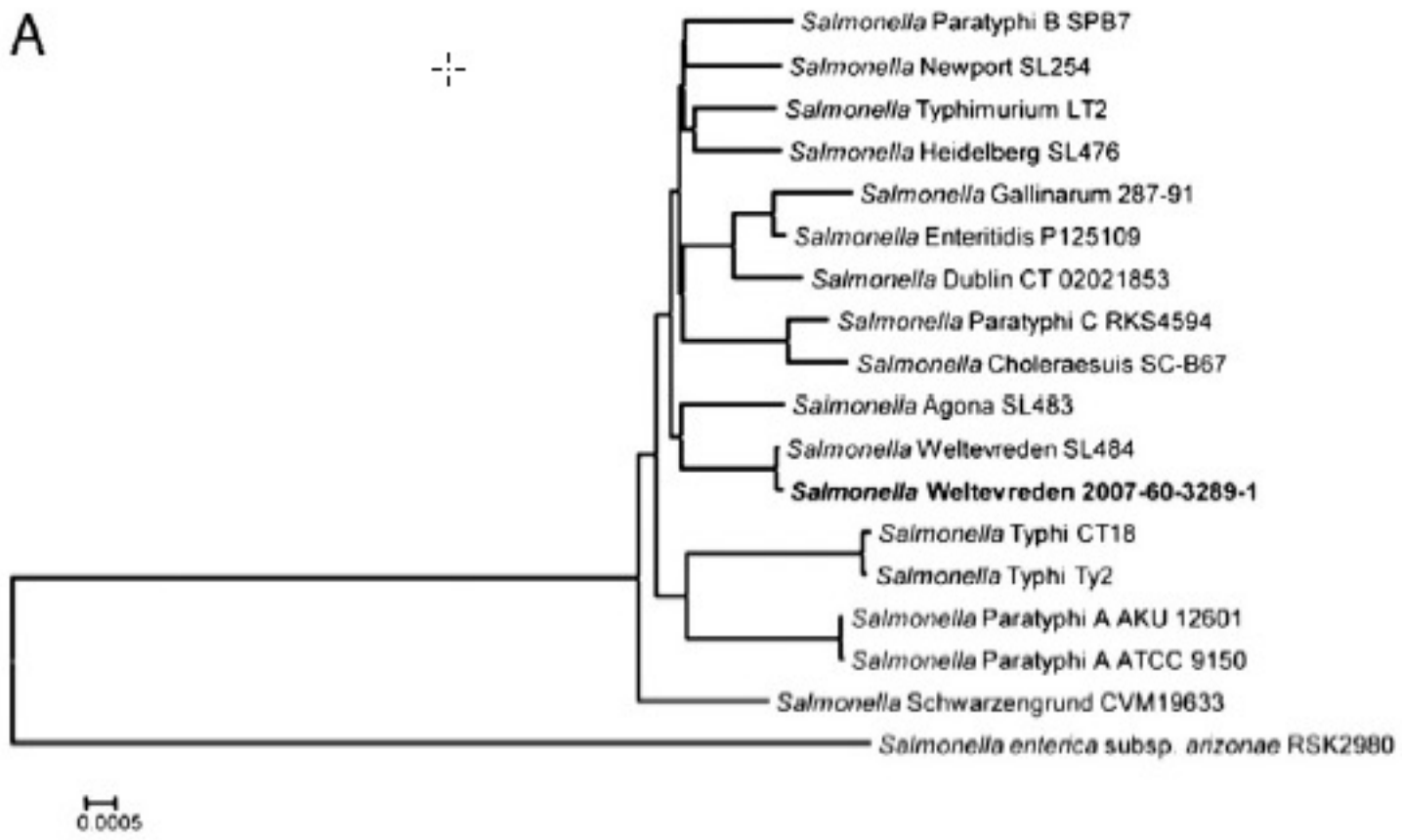


Figure 3.1: Phylogenetic tree generated using the core genes (~2650 coding sequences) of various *Salmonella.* Taken from [238]

S. Agona was found to be genetically the closest serovar to *S.* Weltevreden. The analysis of *S.* Weltevreden 2007-60-3289-1 revealed the presence of additional cluster of genes likely associated with carbohydrate metabolism, suggesting the possibility of survival in alternative habitats [238]. Indeed, the average *S.* Weltevreden genome is slightly bigger than the average *Salmonella* genome. The presence of many of the major *Salmonella* SPIs was confirmed in the Scandinavian isolate. Interestingly comparative analysis between SL484 and 2007-60-3289-1 showed high genetic similarity despite the different source and geographic location of the isolates [238].

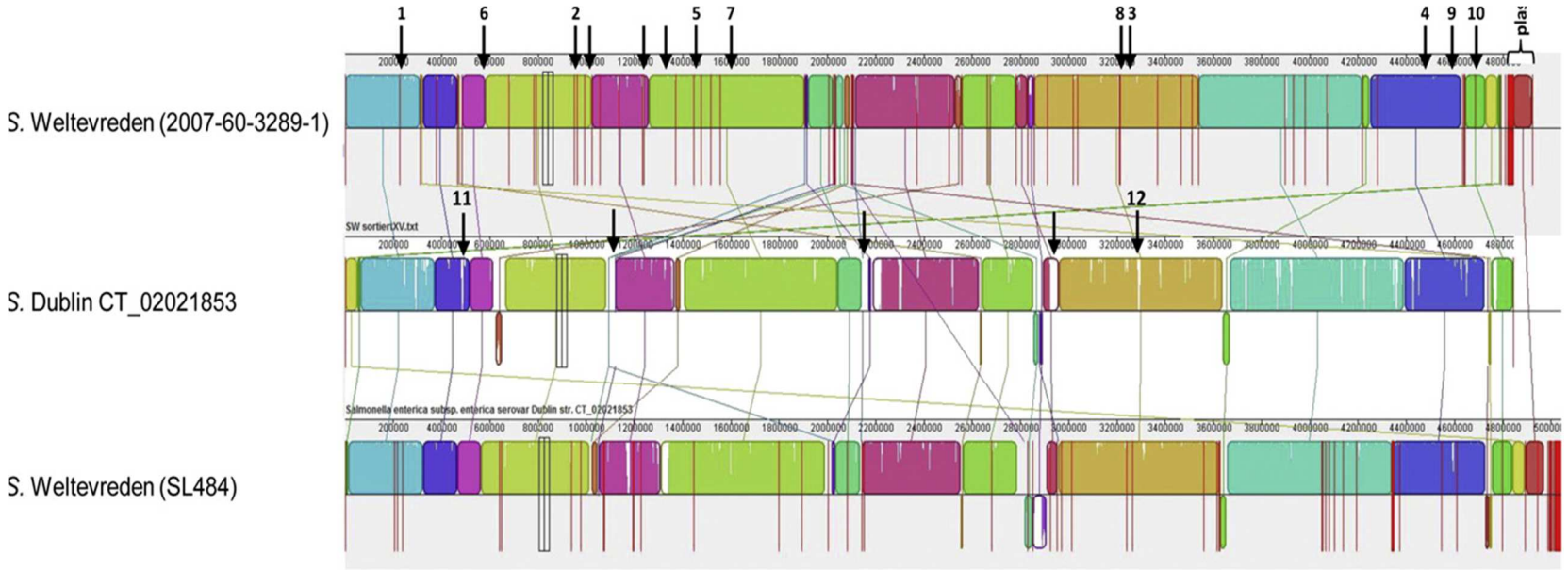


Figure 3.2: Mauve progressive alignment of the draft genomes of *S*. Weltevreden 2007-60-3289-1, S. Dublin CT_02021853 and *S*. Weltevreden SL484

S. Dublin was used as a comparator as this serovar has a similar genome size and shares many features with *S*. Weltevreden 2007-60-3289-1. Black numbered arrows indicate selected features; 1: 1- T6SS1 (Spi 6); 2- T6SS2 (Spi 19); 3- Genomic island 1 (Spi 13); 4- Genomic island II (Spi 13); 5- Genomic island III; 6- Genomic island IV; 7- Genomic island V; 8- Genomic island VI; 9- Myo-inositol utilisation loci; 10- Carbohydrate utilisation cluster; 11- Restriction/modification cluster; 12- Phosphonate metabolism. Regions containing phage-related genes are indicated with a black arrow without a number**. Taken from [238]**

In this chapter, Illumina and PacBio sequencing technology were used to sequence the genomes of a globally distributed collection of 115 *S.* Weltevreden isolates. Subsequently, phylogenetic approaches were exploited to generate a detailed population structure and highlight interesting genetic features of these *S.* Weltevreden.

3.2 Results

3.2.1 The *S.* Weltevreden collection

A collection of 115 *S.* Weltevreden isolates from 18 countries was compiled through collaborative efforts involving Dr. Stephen Baker (Ho Chi Minh City, Vietnam) and Dr. Francois-Xavier Weill (Pasteur Institute, France). The isolates were predominantly collected from *S.* Weltevreden endemic area of the South–East Asian region and some West Pacific countries, as well as the Pasteur Institute isolates that represented Francophile countries and travellers. These isolates were from different sources including the environment, food, animal waste, animals, human faeces and blood and they covered a period from 1940 to 2013. The map depicted below shows the geographical distribution of the isolates included in this study.

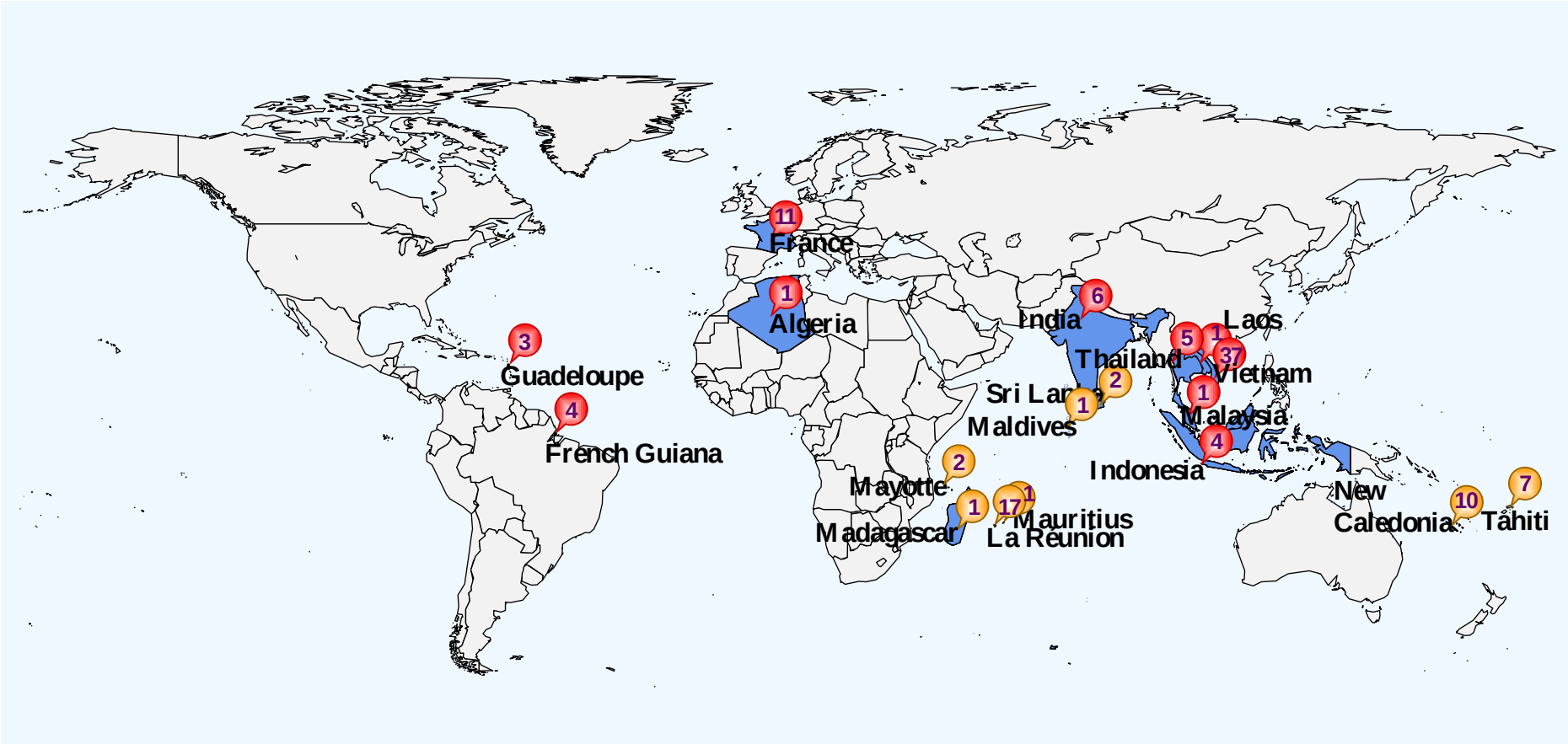


Figure 3.3: Geographical distribution of *S*. Weltevreden isolates included in the study

Colours (red and yellows) represent the phylogenetic clusters and the number of the isolates from each location and will be discussed later in the chapter.

3.2.2 Generation of a *S*. Weltevreden reference genome

In order to gain a better understanding of the genetic architecture of *S*. Weltevreden, DNA from the isolate 10259 obtained from a stool of a diseased Vietnamese child, was sequenced using both the Illumina and PacBio RSII long read sequence platforms. For a full description of the approaches see the Methods section. After manual fixes, a new high quality reference genome for *S*. Weltevrden 10259 was generated which revealed a single contig for the main bacterial chromosome and an additional contig for a large plasmid that is present in many isolates. The chromosome of *S*. Weltevreden 10259 is a single circular molecule of 5,062,936 bases that harbours a 4723 predicted coding DNA sequences (CDSs). The single plasmid is 98,756 bases in length with 98 predicted CDSs. The genome has a G+C content of 52.1%. Putative functions of coding genes were assigned using the Sanger automatic annotation pipeline (Accession number to be provided post submission). Figure 3.4 below represents a map of the genome generated using DNA plotter.

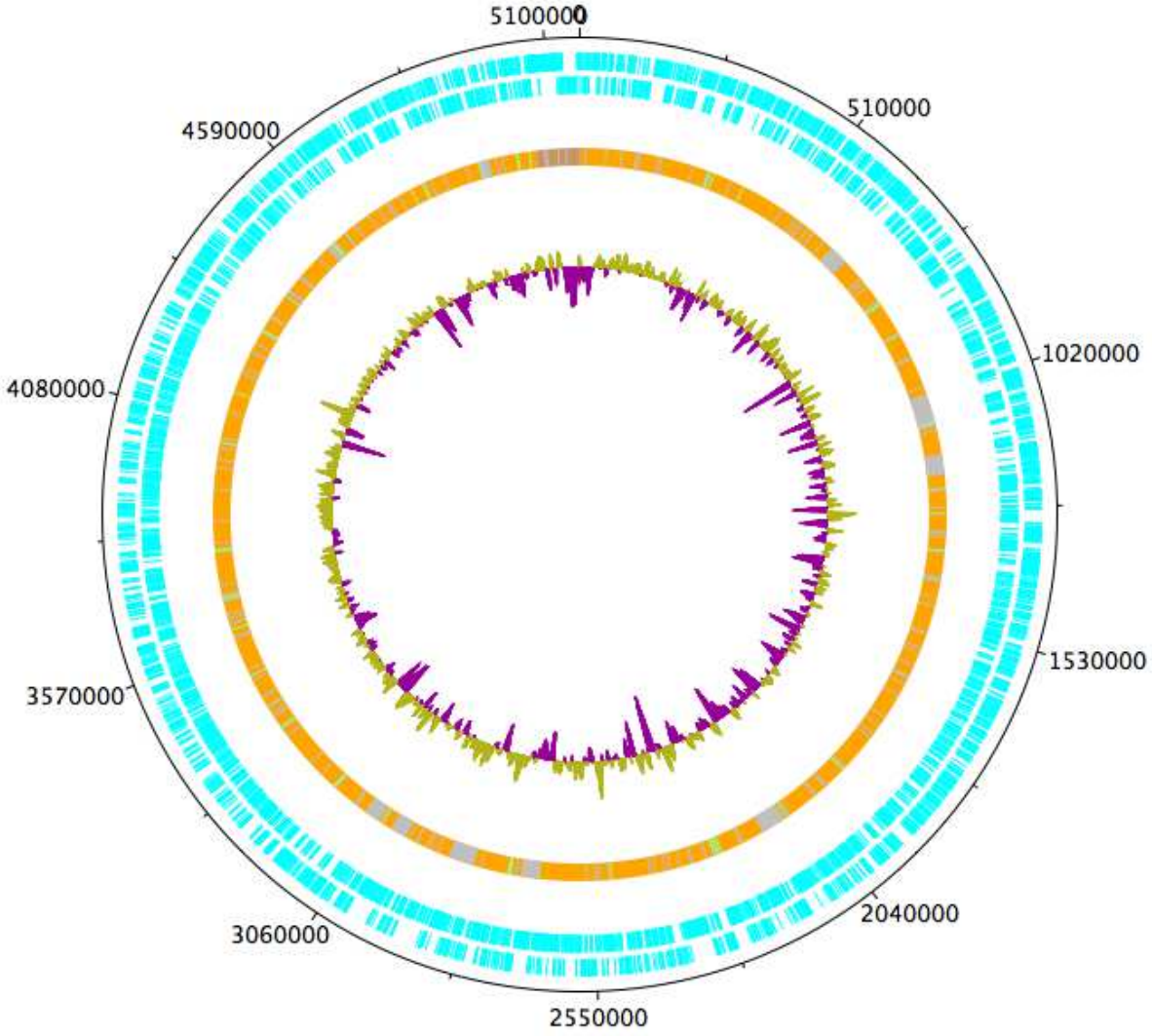


Figure 3.4: DNA plotter diagram of *S*. Weltevreden 10259 genome

The outer black circle designates the genome base positions around the chromosome. The next 2 outer blue circles depict predicted CDSs on both strands. The predominantly orange circle represents the main chromosome with likely horizontally acquired DNA elements. On this circle, grey areas represent non-coding RNA (ncRNA) and the green areas represent tRNA. The inner circle represents the % of GC plot.

3.2.3 Phylogenetic analysis

3.2.3.1 Confirmation of sequence type

DNA from all 115 *S.* Weltevreden isolates were sequenced on the Illumina sequencing platform with an average coverage of ~70 times the genome size. A statistical and logistical analysis of the runs can be found in Appendix 4. To correlate their serovar identity, all 115 sequences together with the other published sequences isolates were run on the reference database http://mlst.warwick.ac.uk/mlst/dbs/Senterica to determine their MLST. The data confirmed all of the isolates were Weltevreden Sequence Type 365 and Table 3.1 below summarises the MLST of 10259 as well as the alleles numbers for each housekeeping gene used to compile the MLST profile.

Isolate	**ST**	**Contamination**	***aroC***	***dnaN***	***hemD***	***hisD***	***purE***	***sucA***	***thrA***
10259	365	none	130	97	25	125	84	9	101

Table 3.1: MLST data for *S.* Weltevreden 10259 reporting alleles numbers.

3.2.3.2 *S.* Weltevreden in the context of other *S. enterica*

S. Weltevrden is a relatively under studied serovar. Indeed, the exact placement of the *S.* Weltevrden serovar in a comprehensive phylogenetic tree based on whole genome sequences inclusive of other *Salmonella* serovars has not been available. Previous limited phylogenetic analysis placed *S.* Weltevrden close to *S.* Agona [240] and *S.* Enteritidis [241] in a separate eburst group [59]. The earlier analysis was compromised by the poor quality of the published reference genome (accession JPIO01) used for those analysis, which included only 1,744 predicted CDSs, a third of the expected number.

To clear up this confusion, the new reference generated in this study was used to compare *S.* Weltevreden with 57 other isolates representative of different *S. enterica*

serovars. A single high quality assembly was chosen for each serovar. Twenty two were high quality finished assemblies and thirty six were fragmented assemblies (see Table 3.2). A pan genome was constructed with *S.* Weltevrden and the 57 other serovars using Roary and a core genome of 2,572 genes was identified, representing 2,382,319 bps, with SNPs at 150,074 positions. The size of the core genome is in line with previously published work. This data was used to create a phylogenetic tree using RAxML with 100 bootstraps. The nearest serovar phylogenetically is *S.* Elizabethville with a difference of 11,989 bases in the core genes of the representative isolates with 0.5% variation. This observation is also supported by the similarities in their serology (*S.* Weltevreden is O:3, O:10 or 15 and r, z_6 positive and *S.* Elizabethville is O:3, O:10 and r, 1,7 positive). *S.* Agona, which has already been reported to be genetically close to *S.* Weltevreden, also mapped closely to *S.* Weltevreden in this independent analysis with 20,916 bp differences in the core genes representing a variation of 0.877%.

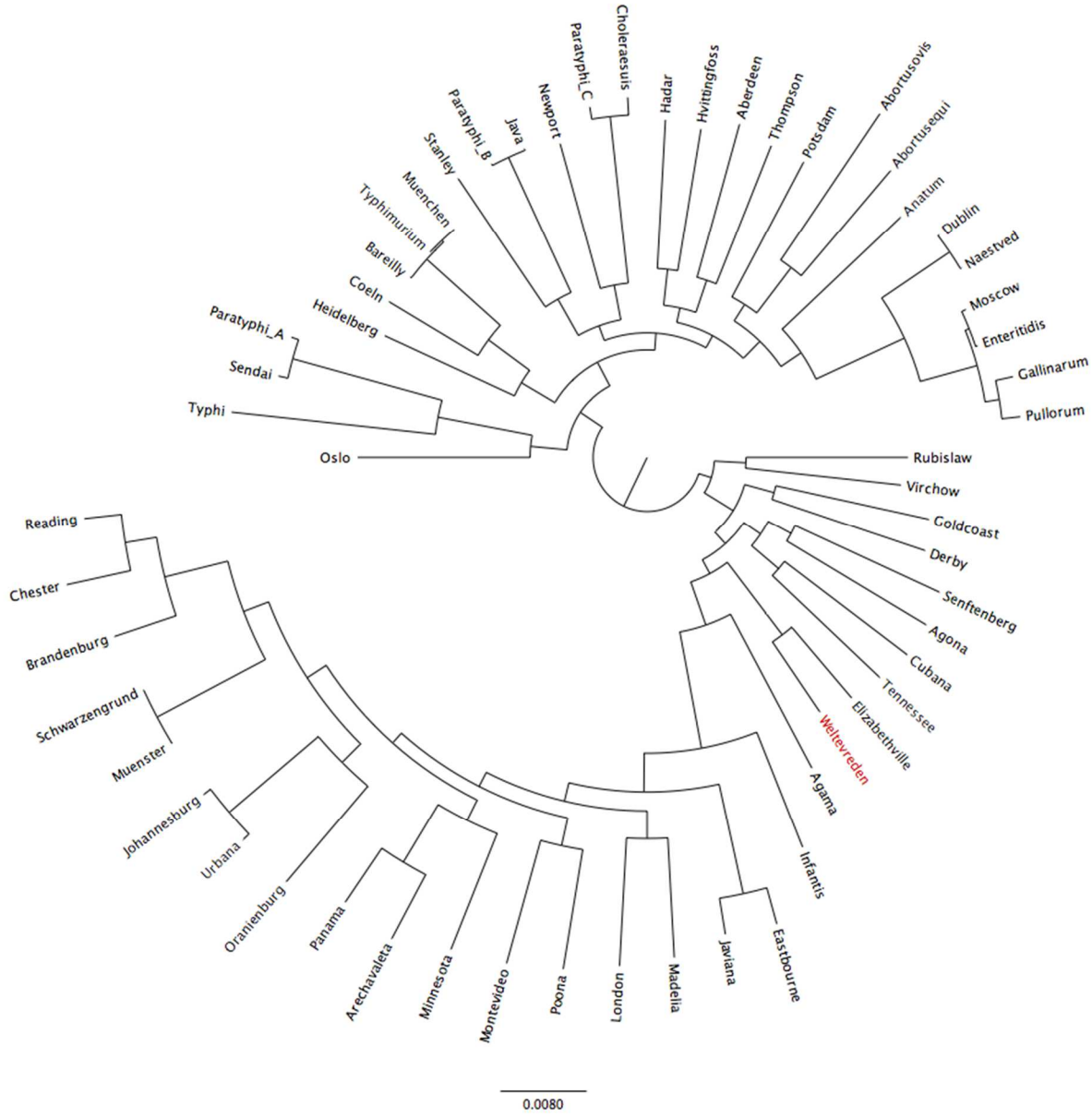


Figure 3.5: Maximum likelihood tree comparing *S.* Weltevreden with selected *S. enterica* serovars

2,572 core genes representing 2,382,319 bps were used to build the tree with SNPs at 150,074 positions.

Below is a table listing the different serovars used to build the global tree.

Serovar	Accession	Type
Aberdeen	ERS179732	Illumina
Abortusequi	ERS179705	Illumina
Abortusovis	ERS179728	Illumina
Agama	ERS037945	Illumina
Agona	CP006876	Complete Reference
Anatum	ERS179686	Illumina
Arechavaleta	ERS016018	Illumina
Bareilly	ERS218079	Illumina
Brandenburg	ERS179748	Illumina
Chester	ERS015989	Illumina
Choleraesuis	CM001062	Complete Reference
Coeln	ERS218091	Illumina
Cubana	CP006055	Complete Reference
Derby	ERS179704	Illumina
Dublin	CM001151	Complete Reference
Eastbourne	ERS179718	Illumina
Elizabethville	ERS394435	Illumina
Enteritidis	Unpublished	PacBio
Gallinarum	CM001153	Complete Reference
Goldcoast	ERS530430	PacBio
Hadar	ERS004922	Illumina
Heidelberg	CP003416	Complete Reference
Hvittingfoss	ERS179726	Illumina
Infantis	CM001274	Complete Reference
Java	ERS207735	Illumina
Javiana	CP004026	Complete Reference
Johannesburg	ERS015996	Illumina
London	ERS179679	Illumina
Madelia	ERS743095	PacBio
Minnesota	ERS015985	Illumina
Montevideo	ERS016005	Illumina
Moscow	ERS179753	Illumina
Muenchen	ERS218081	Illumina
Muenster	ERS016008	Illumina
Naestved	ERS400249	Illumina
Newport	CP001113	Complete Reference
Oranienburg	ERS743094	PacBio
Oslo	ERS179729	Illumina
Panama	ERS016015	Illumina
Paratyphi_A	FM200053	Complete Reference
Paratyphi_B	CP000886	Complete Reference
Paratyphi_C	CP000857	Complete Reference

Poona	ERS668408	Illumina
Potsdam	ERS179673	Illumina
Pullorum	Unpublished	Complete
Reading	ERS179721	Illumina
Rubislaw	ERS218070	Illumina
Schwarzengrund	CP001125	Complete Reference
Sendai	ERS179752	Illumina
Senftenberg	ERS451416	PacBio
Stanley	ERS179745	Illumina
Tennessee	ERS218098	Illumina
Thompson	ERS179680	Illumina
Typhi	AL513382	Complete Reference
Typhimurium	FQ312003	Complete Reference
Urbana	ERS015987	Illumina
Virchow	ERS668352	Illumina
Weltevreden	Unpublished	PacBio

Table 3.2: List of *S. enterica* serovars and Isolates used for comparison and their accession numbers where available

Assemblies fall into 3 general categories, fragmented short read assemblies using Illumina, high quality long read assemblies using PacBio and high quality complete assemblies using a variety of technologies and manual finishing.

A simpler tree was subsequently generated that included *S.* Weltevreden isolates SW10259 and SWC2346, *S.* Elizabethville and several other reference *Salmonella* genomes deposited on the NTCT database (Figure 3.6). Here, *S.* Goldcoast is one of the closest related serovars to *S.* Weltevreden. *S* Goldcoast is mostly associated with zoonosis and rarely infects human with exception of a few reported outbreaks.

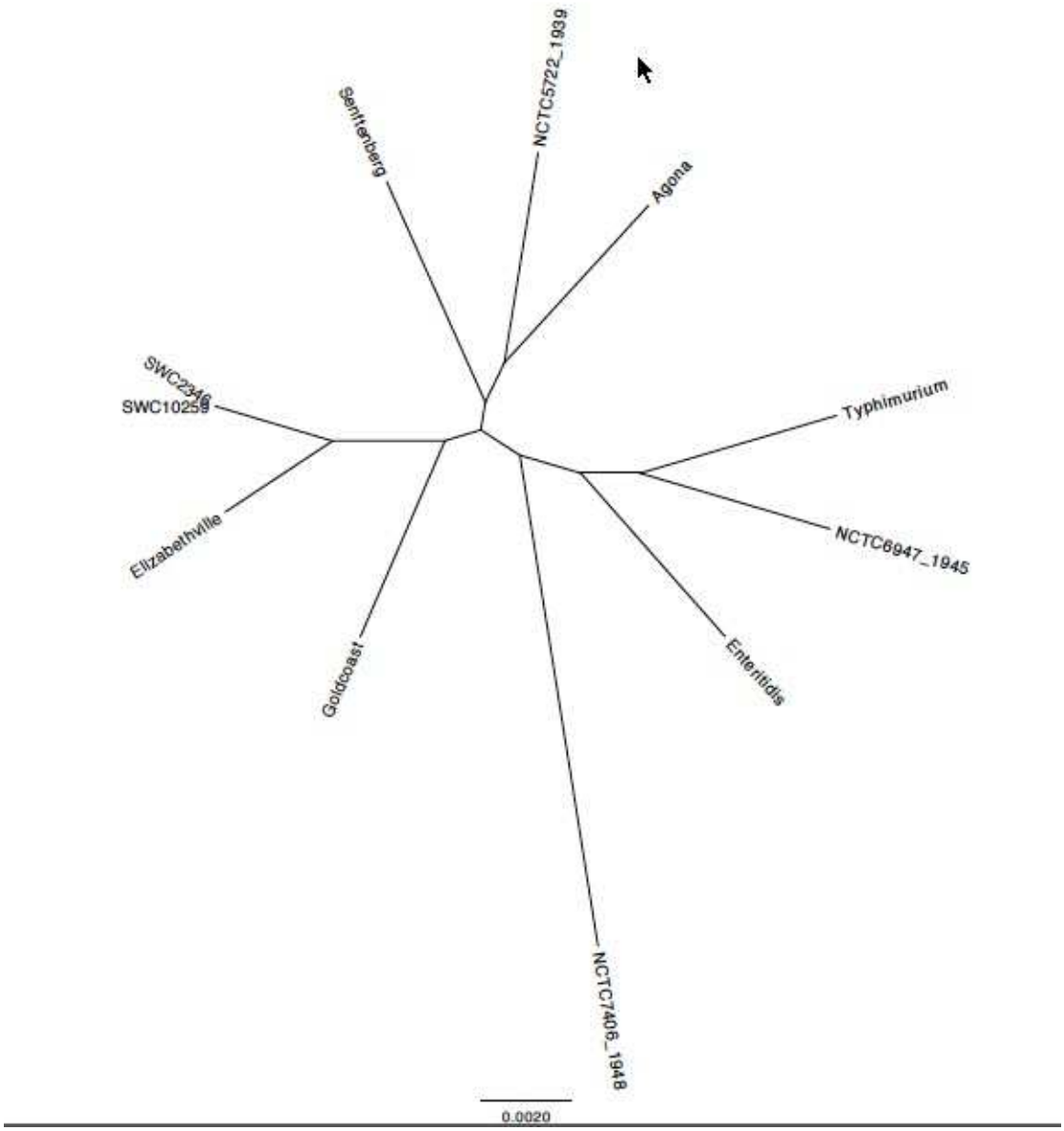


Figure 3.6: Simplified phylogeny of selected *S. enterica* serovars deposited on the NCTC database compared to *S*. Weltevreden.

3.2.4 Genetic diversity of *S*. Weltevreden

To investigate the population structure of *S*. Weltevreden in more detail the genomes of the sequenced *S*. Weltevreden isolates were mapped against the *S*. Weltevreden 10259 reference genome.

3.2.4.1 Chromosome analysis

A multi-FASTA alignment was generated from the aligned samples for the chromosome. SNPs were present at 22,569 positions, but these were subsequently filtered before the generation of the final tree. As the samples were sequenced using Illumina short read technology, repeats longer than the target fragment size (~400 bases) were detected by blasting the reference chromosome against itself and the plasmid sequence against the chromosome. For this analysis, 63 regions greater than 400 bases with a percentage identity of more than 99% were excluded in the multi-FASTA alignment, which resulted in 42 SNPs being excluded. After removal of these long repeats units, Gubbins (a software designed to identify regions of potential recombination that exploits SNP density) was used to limit the effects of recombination on the phylogeny. A total of 218 recombination blocks were identified in the sample set, which reduced the number of core SNPs to 2601. The outcome of this Gubbins-based analysis is shown in Figure 3.7, where the red blocks represent recombination events identified in comparison to the ancestral node and the blue blocks recombination events only present in one isolate.

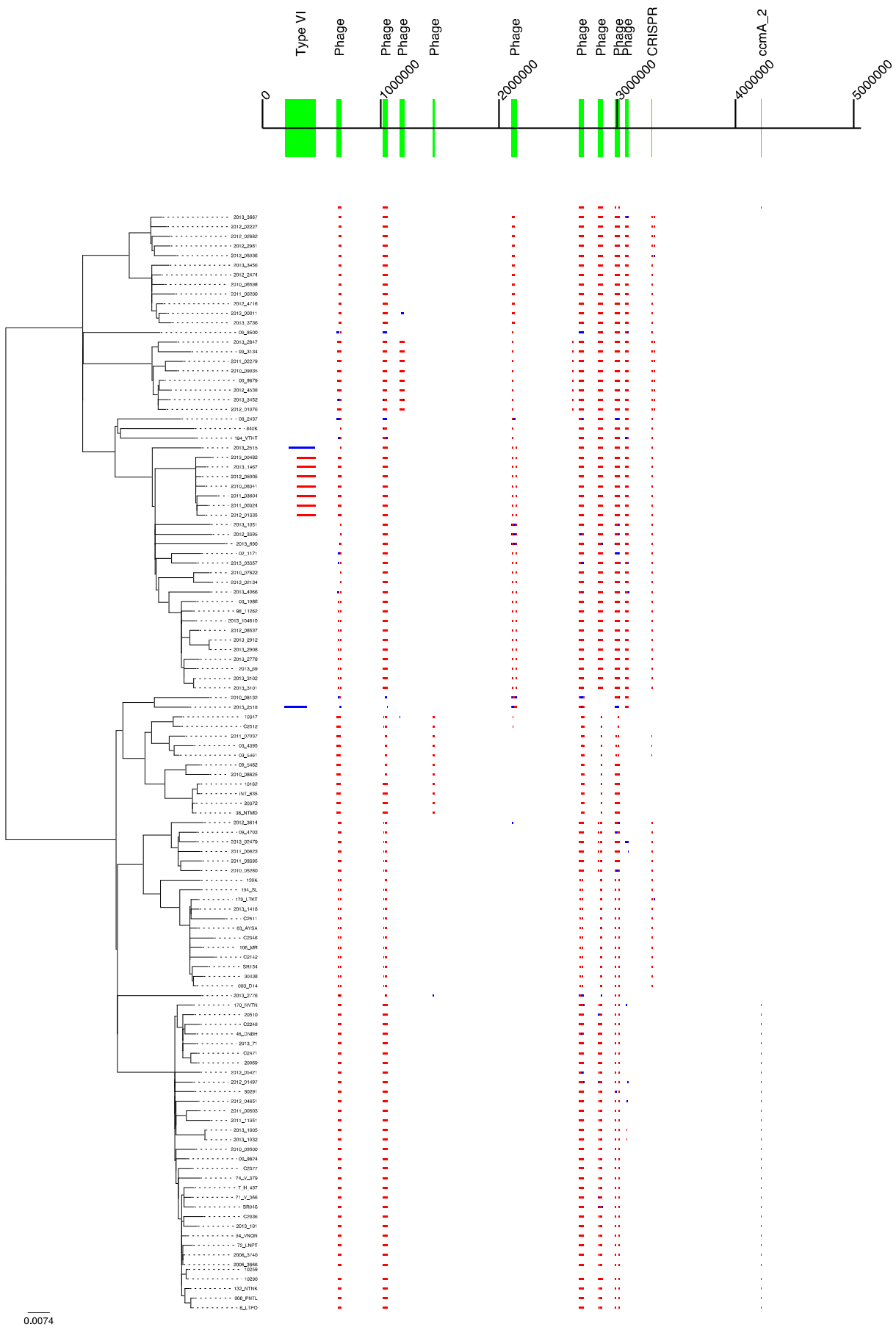

Type VI
Phage
Phage
Phage
Phage
Phage
Phage
Phage
Phage
Phage
CRISPR
cmA_2
0
1000000
2000000
3000000
4000000
5000000
0.0074

Figure 3.7: Gubbins-based analysis representing the recombination events across the tree

The red blocks represent recombination events identified in comparison to the ancestral node and the blue block recombination events only present in one isolate. Recombination regions across the genome are highlighted in green on the reference genome on top.

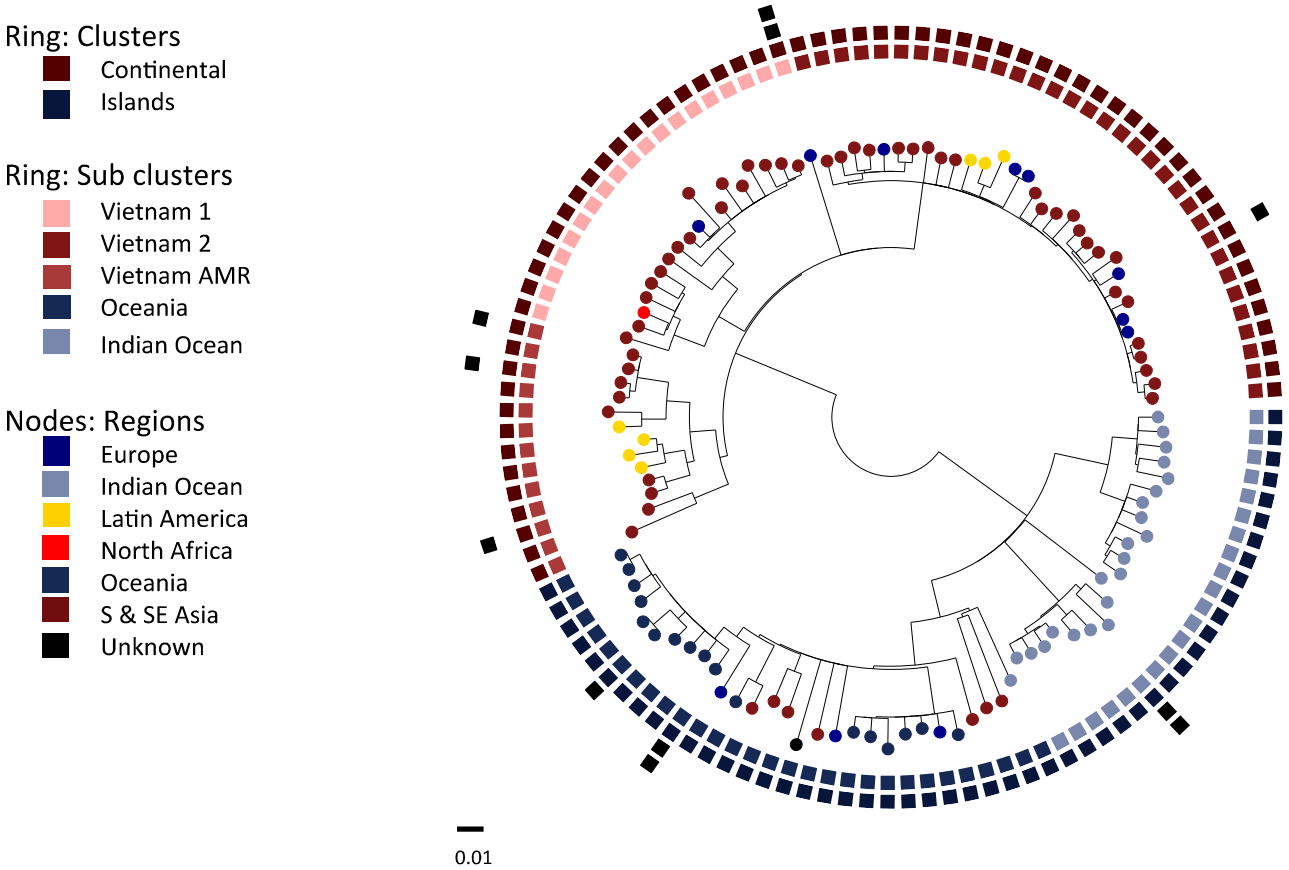
From this refined set of SNPs, a phylogenetic tree was generated with RAxML using 100 bootstraps. The isolates were clustered using BAPS (see methods) identifying 2 primary clusters, which we refer to as the '*Islands Cluster'* and the '*Continental Cluster'* correlating closely in most cases with where the samples were collected. *S.* Weltevreden was definable as a monophyletic serovar sub-dividable into 5 sub-clusters, again correlating closely to where the samples were collected, or their suspected origin (figure 3.8). The *Islands Cluster* contains 2 distinct sub-clusters, one is drawn primarily from islands in the Indian Ocean (*Indian Ocean subcluster*) and the other from islands in Oceania or from nearby South East Asian countries (*Oceania subcluster*). The *Islands Cluster* had a different profile to the *Continental Cluster*. The phylogeny suggests there were independent introductions into different islands and that these subsequently evolved independently. Thus, the phylogeny provides evidence of significant levels of geographical clustering from regional to national. Overall, these data suggest that *S.* Weltevreden slowly evolves within a specific geographical region rather than spreading from one location to another on a frequent basis.

The *Continental Cluster* is dominated by isolates from Vietnam, which reflects the sample bias but this cluster also includes a few isolates from the 'French' islands of the West Indies. There are 3 distinct sub-clusters which capture the circulating lineages, *Vietnam 1*, *Vietnam 2* and *Vietnam resistance*. The *Vietnam resistance* sub-cluster is interesting because out of 14 isolates, 4 have genes linked to antimicrobial resistance and this may be a worrying emerging trend that has not thus far affected *S.* Weltevreden as a whole.

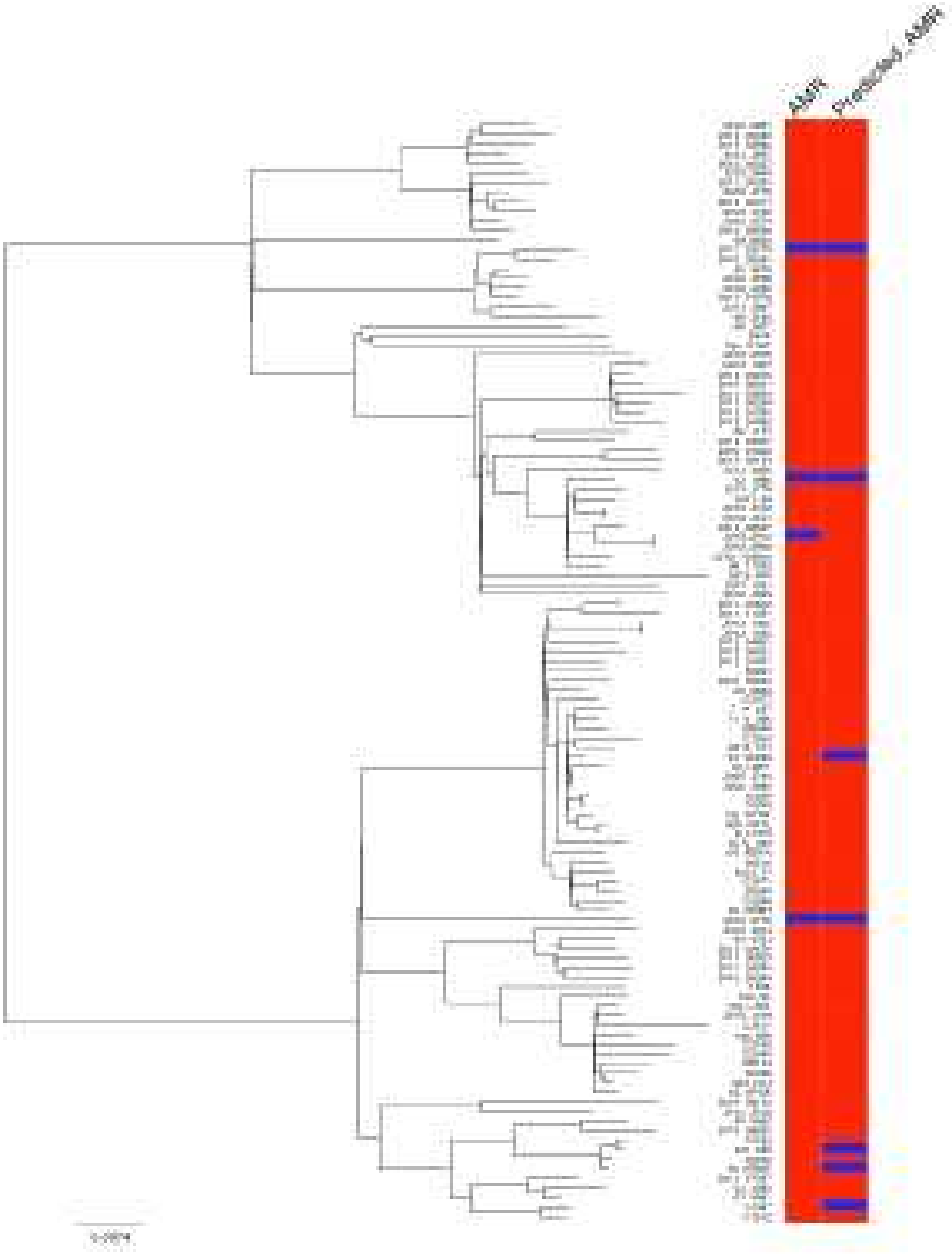
S. Weltevreden isolates obtained from France were scattered throughout the tree without showing any particular clustering. These are likely to be associated, at least in some instances, with international travel. Interestingly, no particular clustering was observed based on the environmental/human source of the isolates. Indeed,

isolates from animal, food, environment and humans were distributed throughout the tree. No evidence of significant temporal clustering was observed.

There are 112 SNPs that can be used to discriminate between the 2 main clusters. These SNPs are the same bases in all isolates of one clusters and different from all isolates outside the cluster (Appendix 3). For instance, in the *aminopeptidase* gene *pepN*, encoding an aminopeptidase, the SNP in position 1,202,670 within the gene is a T in all isolates from the *Continental cluster* and an A in all isolates from the *Island cluster*. These 112 defining SNPs are dispersed evenly throughout the genome with no high density clusters. None of these changes introduce stop codons, so pseudogene formation is not evident.


Ring: Clusters
Continental
Islands
Ring: Sub clusters
Vietnam 1
Vietnam 2
Vietnam AMR
Oceania
Indian Ocean
Nodes: Regions
Europe
Indian Ocean
Latin America
North Africa
Oceania
S & SE Asia
Unknown
0.01

A



B

Figure 3.8: Population Structure of *S.* Weltevreden isolates with key metadata information
Maximum likelihood tree build on 2601 SNPs showing **A**: geographic origin (region), cluster and sub cluster. **B**: antimicrobial profile (laboratory confirmed and predicted); in blue resistant and in red susceptible.

3.2.4.2 Plasmids

A single major plasmid with 98,756 bases in the reference *S.* Weltevreden 10259 is present in 90% of the isolates. It is possible that more of the isolates originally contained this plasmid but it may have been lost from some isolates on storage and culturing as it is missing in a relatively random manner across the tree, particularly from older isolates. There is relatively little variation in the plasmids with only 970 SNPs discriminating plasmids on the tree. If a single potential recombinatorial region is excluded from 7 isolates originating in La Reunion the number of SNPs drops to just 48, which is a variation of 0.048%, indicating a very stable plasmid structure. The plasmid tree structure matches that of the main chromosome, with nearly all the clusters matching identically, indicating that it has evolved with the chromosome. Two of the sub-clusters are interleaved, due to insufficient variation (Figure 3.9). This plasmid shares 99% of similarity with the plasmid pSW82 found in *S.* Weltevreden 2007-60-3289-1 published earlier and contains many classical plasmid genes including toxin and anti-toxin genes, integrase and plasmid maintenance genes (see also Chapter 6). Many isolates contain more than one plasmid, but these are usually medium to small plasmids, which are scattered across the tree.

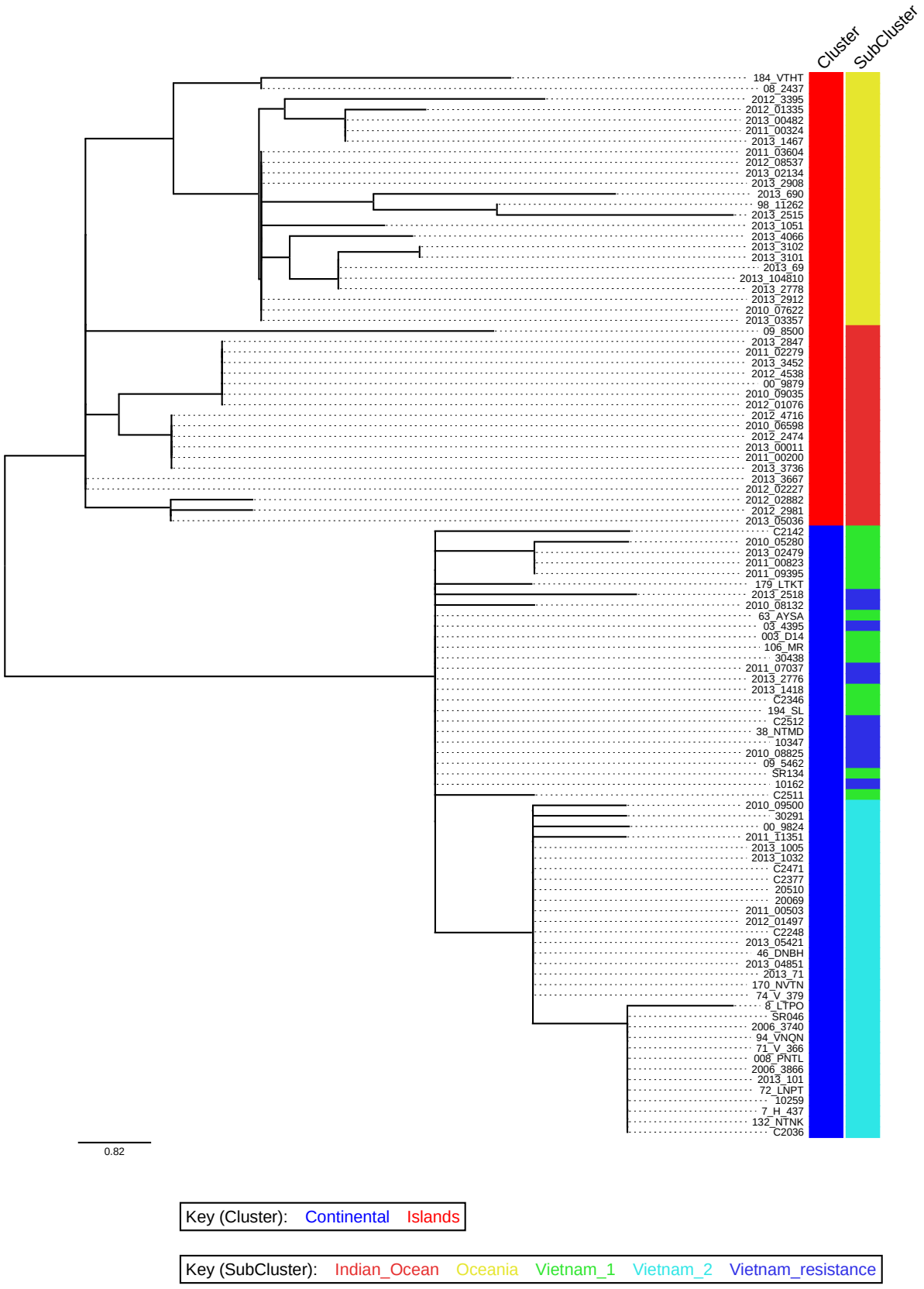


Figure 3.9: Maximum likelihood plasmid tree build on 48 SNPs.

3.2.4.3 Resistome

Antimicrobial resistance has not been frequently reported for *S.* Weltevreden. Indeed, there is very little evident of resistance in the sample set, either phenotypically or through predicted resistance from the genomic data. Known resistance genes were detected in 7 isolates using ARIBA [211], where antimicrobial resistance was conservatively inferred using strict conditions; genes had to be matched throughout the full open reading frame, with 98% identity to the genes known to confer resistance. Isolate 2013_2776, originating from contaminated food in France (potentially imported from South East Asia) was found to have 6 AMR genes, located on a plasmid very similar to the *S.* Heidelberg plasmid pSH111_227 (accession JN983042).

Gene	Accession	10347	38_NTMD	94_VNQN	iNT_635	2011_02279	2013_2776	03_1986
qnrD	FJ228229	R		R				
tetA	AJ517790		R		R			R
qnrS1	AB187515		R		R			
blaTEM30	AJ437107					R		
dfrA1	JQ690541							R
sul3	AJ459418							R
aph3	V00359						R	
oqxA	EU370913						R	
oqxB	EU370913						R	
strA	NC_003384						R	
strB	M96392						R	
tetB	AF326777						R	

Table 3.3: Genes known to confer antimicrobial resistance found in *S.* Weltevreden and the isolates in which they were detected.

3.2.4.3.1 Additional plasmids present in the samples flagged as antimicrobial resistant

A Kraken database was created using all plasmids from RefSeq as well as the *S.* Weltevreden 10259. All of the *de novo* assemblies for the samples which were flagged as having, or potentially having antimicrobial resistance were compared to

the Kraken database, which categorised each contig as a novel plasmid or as related to the *S.* Weltevreden reference chromosome plus plasmid. The novel potential plasmid contigs were extracted for further analysis. This pre-filtering step reduced the size of the input data set by 98%. As the *de novo* assemblies originate from short reads, plasmids are fragmented into multiple pieces. To overcome this issue a nucleotide blast approach was exploited to accurately assign contigs to plasmids and a database was created of all the plasmid sequences in RefSeq. All of the candidate sequences from the *de novo* assemblies were then blasted against these reference plasmid sequences and the novel plasmids were conservatively called (more >95% identity, hits over 10,000 bases in length). Where there were matches to multiple plasmids, a combination of the high values and percentage coverage of the reference plasmid were used.

Table 3.4 lists the 7 isolates with predicted additional plasmids. The equivalents of these plasmids are found in a diverse range of other species and genus. Of the 8 samples with predicted or laboratory confirmed antimicrobial resistance, 7 harboured additional predicted plasmids, equivalents of which have been linked to antimicrobial resistance previously [242-245]. The one phenotypically resistant isolate absent from the list, iNT_635, had too much fragmentation to confidently call a plasmid. However a plasmid related to *pEBG1* is likely present in this isolate (accession NC_025182.1).

Sample	AMR	Predicted AMR	Additional Plasmids	Accession
03_1986	Resistant	Resistant	*S.* Heidelberg plasmid pSH1148_107	NC_019123.1
2013_2776	Resistant	Resistant	*Serratia marcescens* plasmid R478	NC_005211.1
2011_02279	Resistant	Resistant	*E. coli* HUSEC2011 plasmid pHUSEC2011-1	NC_022742.1
38_NTMD	Susceptible	Resistant	*E. coli* strain 09/22a plasmid pEBG1	NC_025182.1
10347	Susceptible	Resistant	*Klebsiella oxytoca* CAV1099 pKPC_CAV1099	NZ_CP011615.1
94_VNQN	Susceptible	Resistant	*K. oxytoca* CAV1099 pKPC_CAV1099	NZ_CP011615.1
2013_2912	Resistant	Susceptible	*Citrobacter freundii* CAV1321 plasmid pCAV1321-135	NZ_CP011610.1

Table 3.4: Isolates likely to harbour antimicrobial resistance-associated plasmids.

3.2.4.4 Accessory genome analysis

The predicted pan genome of the *S.* Weltevreden isolates was created using Roary from annotated *de novo* assemblies. This reference free approach captures much of the sequence diversity, unlike alignment-based approaches, which miss sequences absent in the reference genome. *S.* Weltevreden genomes have ~4500-5000 predicted CDSs, depending on the mobile elements present. A core of 4046 CDSs present in each isolate was identified using this approach against 2572 core CDSs identified early using an alignment based approach on a wider set of serovars. The total accessory genome consisted of 7923 CDSs as show in Figure 3.10.

"Get homologues software" was used to estimate core and pan genome sizes and generate a parse-pan-genome matrix in order to compute and graph the core, cloud, and shell genome compartments [246]. GET_HOMOLOGUES defines these compartments empirically, as follows: core, genes contained in all genomes analysed considered; soft core, genes contained in 95% of the genomes analysed, as described in the study described in [246]; cloud, genes present only in a few of genomes analysed and shell, the remaining genes, present in several genomes.

An average of 15 new predicted CDSs was added to the pan genome with every new isolate (Figure 3.11) and there are an underlying number of unique genes that are only found in one isolate. Some of these gene calls are theoretically due to DNA

contamination and/or mis-assemblies; however the majority appear to be attributable to mobile elements or gene islands. Given the rate of new gene acquisition and the increasing size of the pan genome as seen in Figure 3.12, *S*. Weltevreden appears to have a relatively open pan genome.

Consequently, the size of the potential gene pool from *Salmonella* and in particular *S. enterica*, along with the commonality of a variety of mobile elements, indicates that more sequencing will be required to fully capture the total pan genome of this serovar.

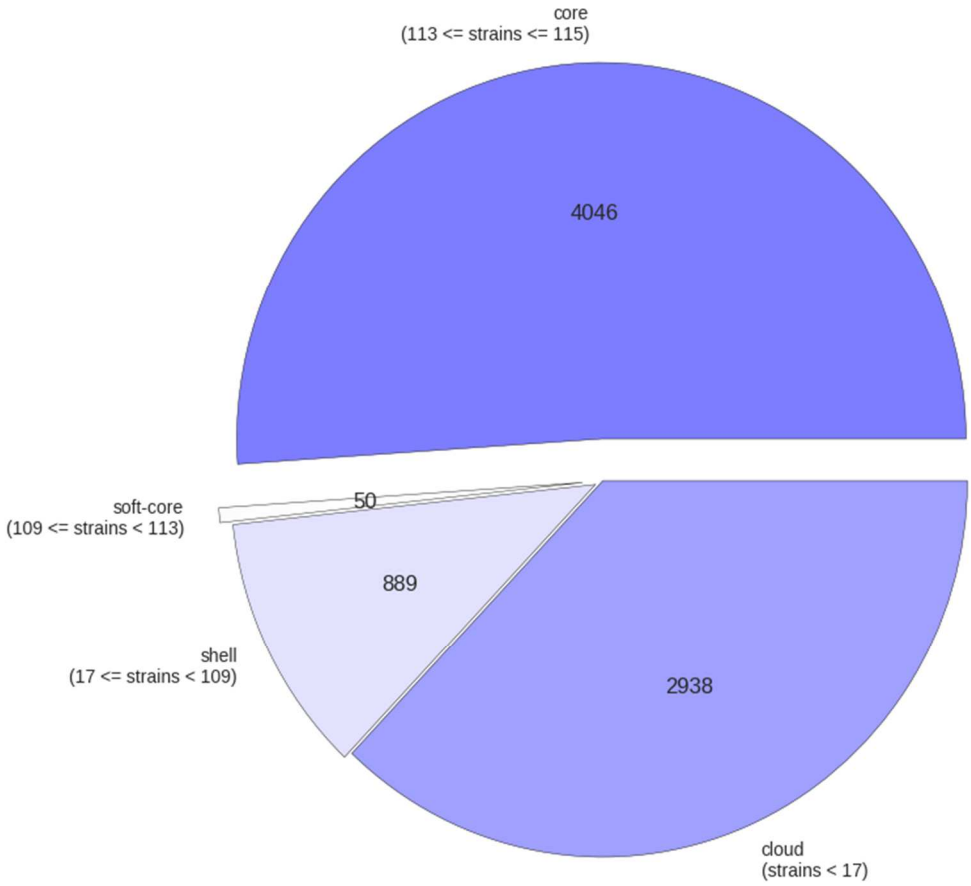


Figure 3.10: Breakdown of the frequency of gene in isolates and in the overall collection of *S.* Weltevreden

Here, the core genome is defined by genes present in 99-100% of isolates, the soft-core by 95-99%, the shell by 15-95% and the cloud by 1-15% as defined in (Contreras-Moreira and Vinuesa, 2013) [246].

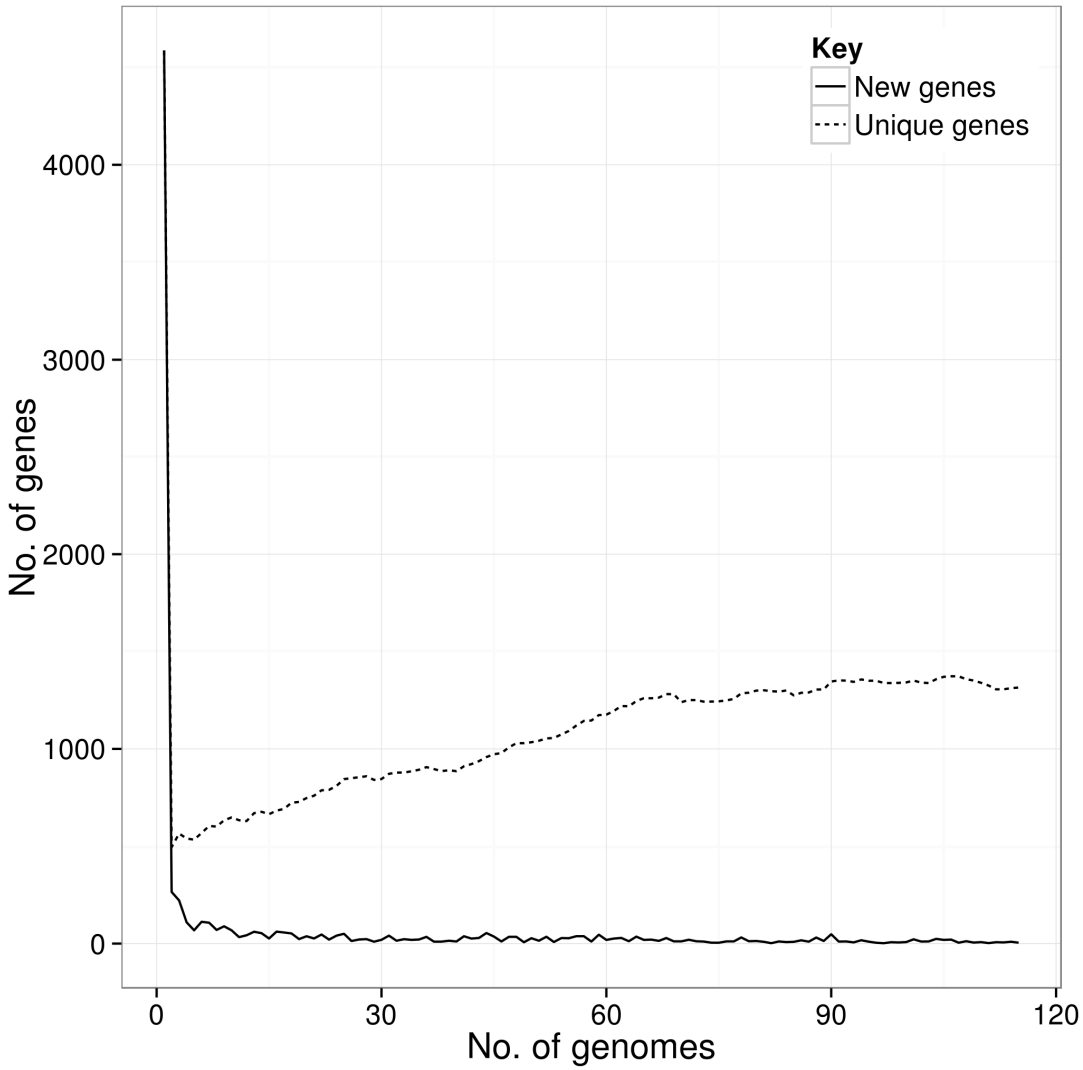


Figure 3.11: Plot showing variance in the number of unique genes found in one isolate only and the number of new genes as genomes are added to the pan genome.

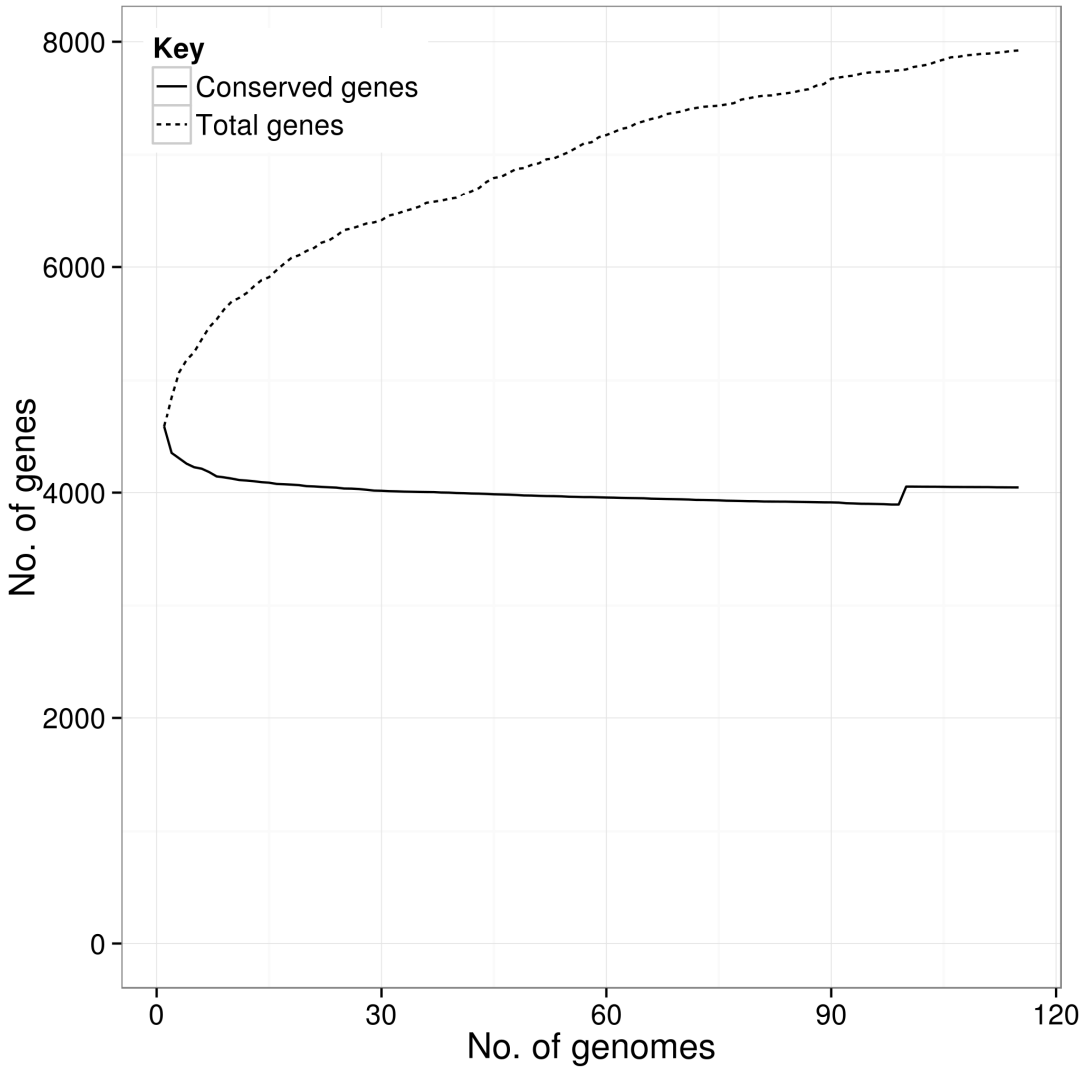


Figure 3.12: Variance in the total number of predicted CDSs (genes) in the pan genome and the of conserved CDSs (99% of isolates) in the core genome as samples are added.

3.2.4.5 Phylogeny of the distribution of phage-like elements

In order to address the phylogeny of genetic elements with phage-like signatures on the tree, the genome sequences of the isolates 10259, C2346, 98_11262 and 99_3134, representing some of the diversity within the tree, were investigated using the website “PHAST”. Several relatively complete phages were identified on each isolates, with an average of 12 phage elements per isolates. Figure 3.13 illustrates the distribution of such phages within the genome of *S*. Weltevreden 10259.

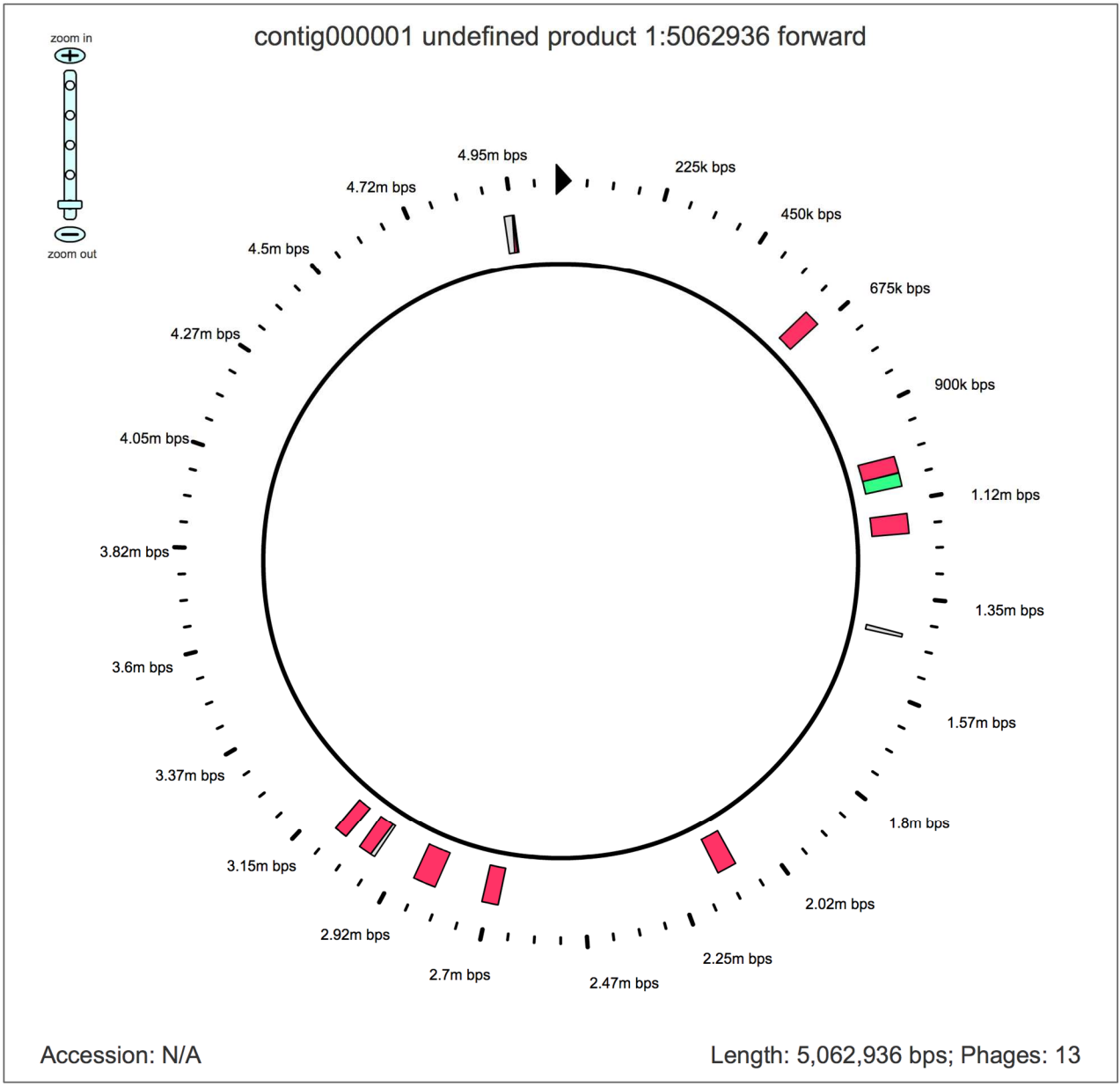


Figure 3.13: Predicted phage distribution across the genome of the *S*. Weltevreden 10259

The outer circle defines the position in the genome; the red blocks represent the larger phage elements found across the genome while the green blocks represent partial phage sequences. Picture generated using Phast

Most phages identified on PHAST were shared by all isolates; these include classical *Salmonella* phages such as Gifsy 1 and 2, Fels 1, as well as entero PsP3 [247]. Figure 3.14 below shows ACT snapshots comparing the 4 genomes at the DNA level. The spaces apparent in each genome likely represent recently acquired DNA. Indeed, more detailed analysis revealed many had signatures of phage or mobile elements and corresponded to regions identified in the PHAST search (Appendix 4).

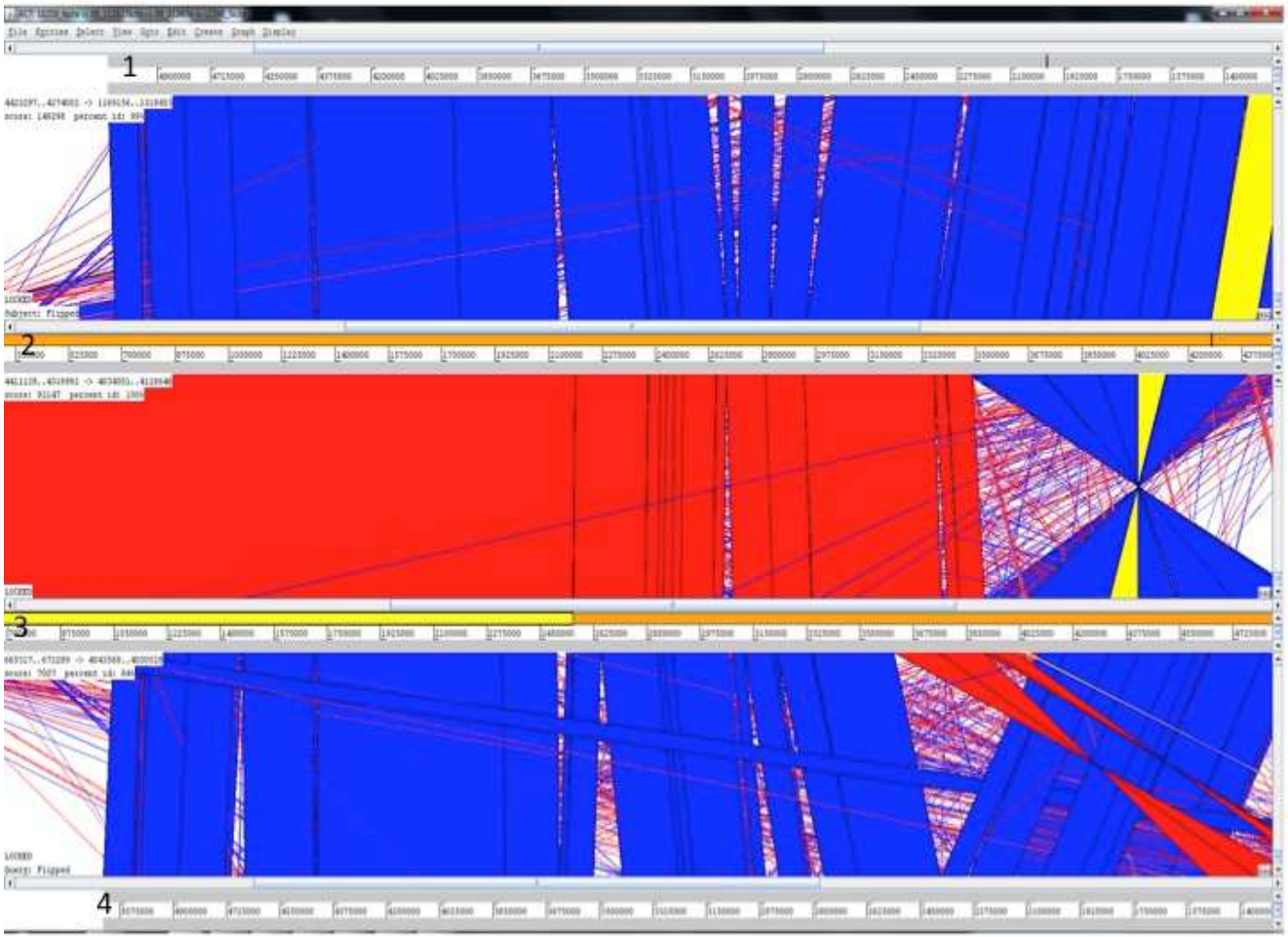


Figure 3.14: Comparative analysis of the genome sequence of *S.* Weltevreden (1) 10259 (2) 98_11262 (3) 99_3134 and (4) C2346

Red, blue and yellow areas represent core genetic elements and the white areas represent potential mobile genetics elements. Analysis performed in ACT

3.3 Discussion

In this study we analysed the genomes of 115 *S.* Weltevreden isolates collected from 18 countries representing the most comprehensive genomic study on *S.* Weltevreden to date involving whole genome sequencing technologies. These analyses revealed that the average *S.* Weltevreden genome is relatively large compare to most *S. enterica* serovars. Indeed, the genomes of all the *S.* Weltevreden isolates were above 5,000000 bps. *S.* Elizabethville was phylogenetically the closest *S. enterica* serovar to *S.* Weltevreden, nicely linking the genotype to the serological phenotype. Indeed, serological test confirmed the high similarities between the 2 serovars with *S.* Weltevreden being O:3, O:10 or 15 and r, z_6 positive and *S.* Elizabethville being O:3, O:10 and r, 1,7 positive. This population structure analysis revealed that *S.* Weltevreden is a monophyletic serovar organised into 2 major clusters of largely continental and island isolates, displaying high level of geographical clustering. There was some evidence of geographical clustering within the sub-phylogeny, suggesting that the *S.* Weltervreden serovar continues to evolve within a specific geographical region rather than frequently spreading from one location to another. There was no obvious temporal clustering within the phylogeny, although this could be because most of our isolates were of relatively recent origin. Importantly, the phylogeny did not correlate with disease type or source of the isolate. For example, it was not possible to distinguish phylogenetically between isolates from diseased patients or controls or even for a specific disease types. Thus, it was not possible to link specific genotype to any disease syndrome.

The two main clusters can be discriminated between using the 112 SNPs (Appendix 3). These SNPs will enable the design of specific probes that could be used in SNP or PCR analysis to discriminate between isolates and allocate novel *S.* Weltevreden to the appropriate cluster. For example, primers flanking selected regions containing the defining SNPs could be used in diagnostic and epidemiology analysis to assign isolates to particular phylogenetic clusters or even to likely country/region of origin. This could be especially useful for travellers returning from multiple destinations to trace down the region of contamination and to further understand the burden of *S.* Weltevreden infection across the globe.

Interestingly, despite the diversity of geographical origin 104 isolates out of 115 possessed a highly related plasmid to that in the reference *S.* Weltevreden 10259, as illustrated by a plasmid tree built with 48 SNPs. Indeed, removal of a single recombination region found in 7 isolates from the same area brought the number of plasmid-associated SNPs down from 970 to 48. This plasmid is more than 99% identical to the plasmid pSW82 found in *S.* Weltevreden 2007-60-3289-1.

Antibiotic resistance was not a very common phenotype within our *S.* Weltevreden collection and genome analysis for the presence of antibiotic resistance signatures within the genomes of our collection confirmed the laboratory findings, with just 7 isolates out of the whole set displaying antimicrobial resistance genes. Further analysis suggested that these resistant isolates harboured novel plasmids of types previously described in other bacteria. Thus, *S.* Weltevreden, although largely still susceptible to antibiotics, has the potential to acquire multiple antibiotic resistance and any trends in this direction should be carefully monitored in the future.

A total of 4046 core predicted genes present in each isolate were identified using reference free accessory genome analysis. *S.* Weltevreden appears to have a relatively open pan genome based on the rate of new gene acquisition and the size of the pan genome. However, further analysis will be required for a comprehensive description of the *S.* Weltevreden accessory genome. Initial work showed that phages and mobile elements varied depending on the isolate. Well characterised *Salmonella* phages were present as well as more novel phage types. Thus, in common with other *S. enterica*, phage and other mobile elements are a key driver of diversity, suggesting that the serovar is undergoing rapid and continuous evolution.

The ability to generate DNA sequence and to construct accurate phylogeny will facilitate further epidemiological analyses and functional genomic work designed to link phenotype to genotype. The initial steps in this direction described in this Chapter benefited greatly from the generation of an accurate reference genome *S.* Weltevreden 10259 that will be made available to the community. This genome can provide a basis for further functional genomic work, including mutagenesis, RNA-seq and proteomics.

4 Phenotypic characterisation of *S.* Weltevreden

4.1 Introduction

The Kauffman-White scheme has been used for decades to classify *Salmonella* into serovars based upon serological analysis. This phenotypic approach has proved to be invaluable in public health terms for epidemiological and clinical work, although the level of resolution limits its utility. Work described in the previous chapter defined the levels of genetic diversity within a collection of *S.* Weltevreden isolates at a whole genome level using phylogenomic approaches. These studies defined *S.* Weltevreden as a monophylectic group that could be divided into sub-clades with distinct genetic structures. Our interest in this serovar was stimulated in part by increasing reports of an association of *S.* Weltevreden with clinical disease in different geographical regions. For example, our collaborators working at the Oxford University Clinical Centre in Ho Chi Minh City, Vietnam were isolating *S.* Weltevreden from both diseased and control individuals (see Appendix 1). Most of the clinical disease was associated with gastroenteritis and there was an indirect link with marine food sources (Dr. S. Baker, personal communication). Additionally, *S.* Weltevreden is emerging as one of the most frequently isolated serovars in clinical salmonellosis cases from other regions.

Despite an increasing effort to genetically characterise *S.* Weltevreden, little to no phenotypic data is currently available in the published literature Thus, there is a lack of knowledge and a clear understanding of the mechanisms of microbial pathogenesis associated with this serovar. Such information would be of value for designing approaches to prevent and tackle disease. Thus with an aim to gain more insights into the pathogenesis and host response to infection stimulated by *S.* Weltevreden, we embarked on a series of experiments designed to phenotype selected isolates representing the diversity of the phylogenetic tree. These isolates

were tested in various experimental settings to define characteristics using isolates representative of specific genetic clusters (where isolates were available) with the overarching goal of linking the genotype to the phenotype. A comparative analysis of isolates of the *S.* Weltevreden and *S.* Typhimurium serovars was performed in each experiment in order to capture the key differing features between these 2 groupings.

For the experiments described in this chapter, *S.* Typhimurium SL1344 was used as a control as this isolate has been used extensively in laboratories around the world and in various *in-vitro* and *in-vivo* models of infection [248, 249]. For example, as a mouse pathogen, *S.* Typhimurium SL1344 has been used in both the typhoid [250] and streptomycin treated gastroenteritis (colitis) mouse model [160]. Thus, here we exploit the extensive genetic and phenotypic data already available on *S.* Typhimurium SL1344 to characterise and compare *S.* Weltevreden isolates to this more common nontyphoidal *Salmonella*.

4.2 Results

4.2.1 Microbial characterisation and confirmation of serotype

S. Weltervreden isolates SW C2346, SW 10259, SW98_11262 and SW99_3134 were selected for use in phenotypic assays. Their positions within the *S.* Weldervreden phylogeny are shown in Figure 3.8. As a step towards validating that the isolates were phenotypically *S.* Typhimurium or *S.* Weltervreden they were propagated on L-agar and characterised for microbial growth and serotyped using reference serotyping sera. Initially, an agglutination test was performed on *S.* Typhimurium SL1344 using O4 and 05 sera, according to the Kauffman-White classification for *S.* Typhimurium. O10 typing sera were used as a negative control. The agglutination data for *S.* Typhimurium SL1344 were consistent with the Kauffman-White serological classification. In order to further validate that this isolate was SL1344 and not the *aroA* mutant derivative SL3261, which is also frequently used in the laboratory, the presence of the *aroA* gene was confirmed by colony PCR by amplifying the *aroA* region using primers specific for *S.*

Typhimurium. The isolate was also able to grow on media lacking aromatic supplements. These assays confirmed that the isolate did not harbour an *aroA* mutation.

Similarly, to validate the serotype of the *S.* Weltevreden isolates, agglutination tests were performed on all isolates. Based on the Kauffman-White classification, *S.* Weltevreden is O3 positive, O10 or O15 positive for O antigen and, R and Z_6 Positive for phase 1 and 2 H antigens respectively. Here, O4 typing sera was used as a negative control and the tests were performed as described in the methods section. Additionally, all isolates were Vi-negative. Table 4.1 below summarise the results.

Strain	**Somatic antigen**				**Flagella antigen**		**Virulence antigen**
	03	**04**	**010**	**015**	**R**	**Z_6**	**Vi**
SW C2346	+++	-	+++	-	+	+++	-
SW 10259	++	-	-	+++	++	-	-
SW 98_11262	++++	-	++++	-	+	+	-
SW 99_3134	+++	-	++++	-	+	+	-

Table 4.1: Sera agglutination results summary

-: no agglutination observed, +: low agglutination, ++: mild agglutination, +++: strong agglutination, ++++: very strong agglutination.

All isolates were strongly positive for the core-typing antigen O3 whereas isolates SW C2346, SW98_11262 and SW99_3134 were additionally positive for O10. Isolate SW 10259 was positive for O15 but not O10. All isolates were negative for the control O4 typing sera. All 4 isolates were positive for H_R and isolates SW C2346, SW98_11262 and SW99_3134 were positive for the Phase 2 antigen Z6. Again, SW 10259 was distinct in that it did not react with Z6 sera. This may be expected as Phase 2 antigens are not always expressed. The data retrieved from the

tests were generally consistent with the Kauffman-White scheme and taken together with the phylogenetic data the isolate fall into the *S.* Weltervreden serovar grouping.

4.2.2 Bacterial growth *in-vitro*

To ensure that any further phenotypic difference observed between the isolates was not due to any general differences in their growth rate this was assessed in LB medium over the course of 24 hours at 37° C. The results are shown in Figure 4.1. All isolates grew with a similar doubling time in this medium.

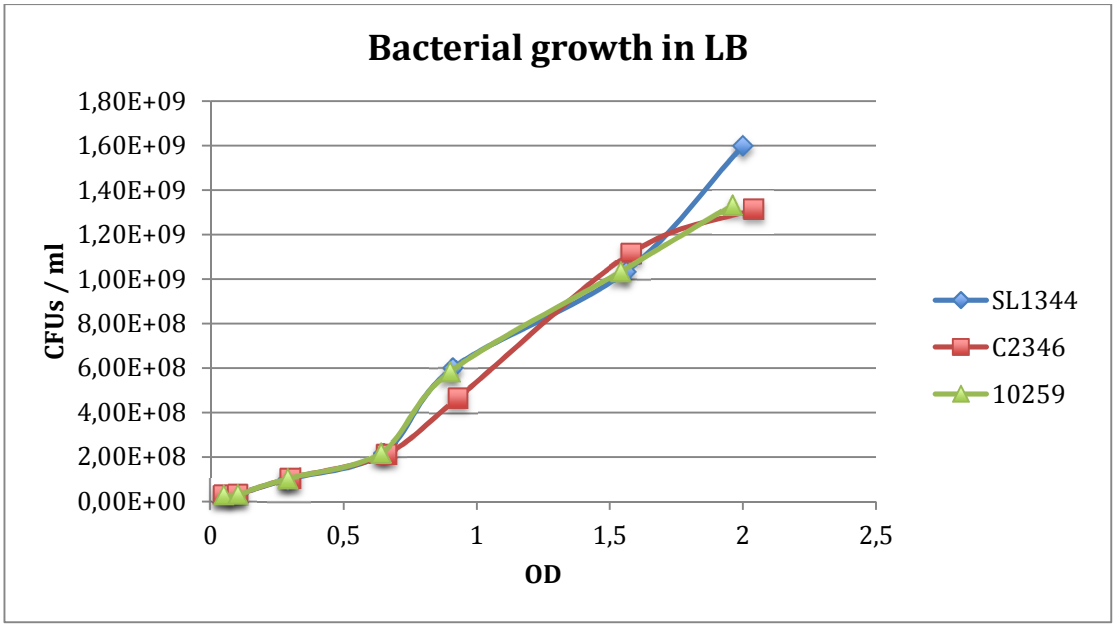


Figure 4.1: Standarisation curve reporting the number of colony forming units (CFUs) per OD600 in LB medium for each bacterial isolate

Biological triplicates were used while performing this experiment.

4.2.3 Metabolic profiling using the Biolog Phenotype Microarray system

To gain more insight into the phenotypic diversity of *S.* Weltevreden compared to *S.* Typhimurium, 2 of the *S.* Weltevreden isolates, SW C2346 and SW 10259 were tested alongside *S.* Typhimurium SL1344 for metabolic activities using Biolog Phenotype Microarray™ plates and any metabolic differences were scored and analysed using the Biolog OPM data analysis platform [251]. Plates PM1 and PM2

(covering different carbon sources), plate PM3 (Nitrogen sources), plate PM4 (phosphorus and sulfur sources), plate PM9 (osmolytes) and plate PM10 (pH) were used in the assays, which were repeated several times. Overall the different isolates had a consistent metabolic profile with most individual assays reporting similar results. Both *S.* Weltervreden gave indistinguishable metabolic profiles. The key metabolic differences between *S.* Weltervreden and *S.* Typhimurium are shown in Figure 4.2.

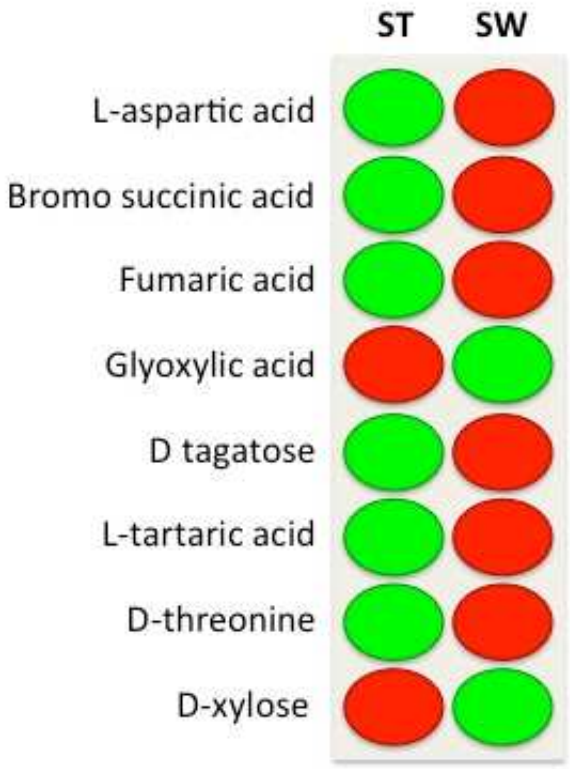


Figure 4.2:Carbon source utilisation microarray

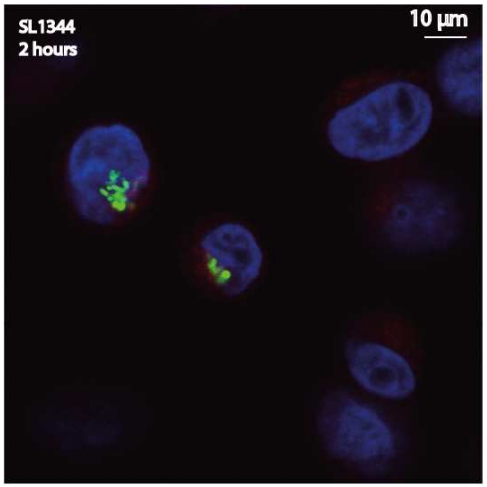
The carbon sources differentially utilised by *S.* Weltevreden (SW) and *S.* Typhimurium SL1344 (ST). The green circles represent the dominantly utilised energy sources and red circles represent the less favoured sources for each serovars. This data is representative of multiple biological replicates.

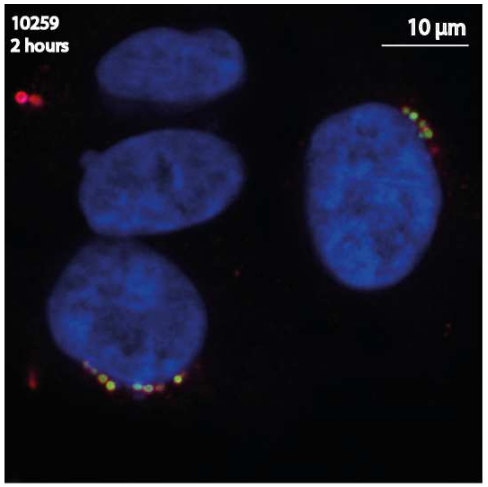
The results confirmed that both *S.* Weltevreden isolates have the capacity to exploit similar carbon sources (Appendix 5). In contrast, notable difference in carbon choices was observed between *S.* Typhimurium SL1344 and the *S.* Weltevreden isolates. D–xylose has been associated with amino sugar and nucleotide sugar

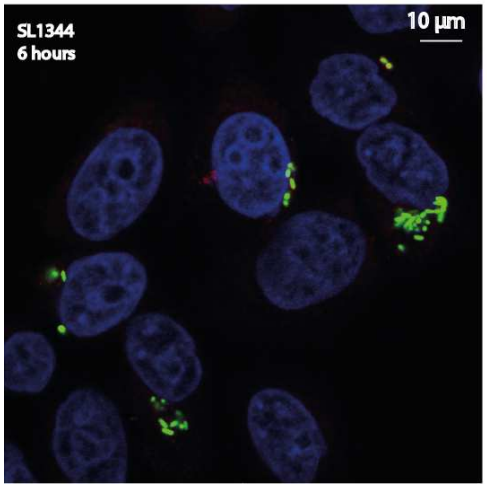
metabolism, pentose and glucoronate inter-conversion as well as starch and sucrose metabolism [252, 253]. Glyoxylic acid is associated with purine and amino acid metabolism, carbon fixation pathways, biosynthesis of secondary metabolites and microbial metabolism in diverse environments.

4.2.4 Invasion into Hep 2 cells

Eukaryotic host cells growing in culture have been used extensively to monitor cellular interactions between host and microbe. A classical method involves the exposure of cell lines growing *in-vitro* to different numbers of either wild type or mutant pathogens followed by microscopic or microbiological observations. *Salmonella* have the ability to both adhere to and invade cultured cells, including macrophage and epithelial cell lines. Exposure to the antibiotic gentamicin, which is a poor killer of intracellular bacteria, has been routinely used as a method for estimating invasion levels. Consequently, cultured Hep 2 cells were exposed independently to either *S.* Typhimurium SL1344(pSsaG) or one of *S.* Weltevreden SW C2346(pSsaG), SW 10259(pSsaG), SW98_11262(pSsaG) or SW 99_3134(pSsaG) at a multiplicity of infection (MOI) of ~50 bacteria per cell. Plasmid pSsaG directs the expression of GFP from the SPI-2-associated *ssaG* promoter. Thus, host bacteria only become green when they have established a SCV (Figure 4.3).


SL1344
2 hours
10 µm


10259
2 hours
10 µm


SL1344
6 hours
10 µm

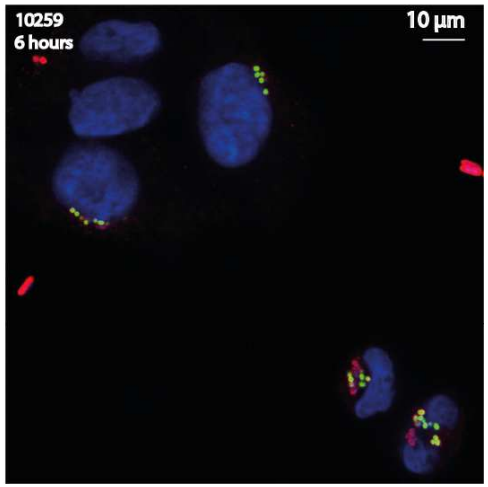

10259
6 hours
10 µm

Figure 4.3: Confocal microscopy of *S*. Typhimurium SL1344 and *S*. Weltevreden in Hep 2 cells, 2 and 6 hours post exposure

Cell nuclei were stained with DAPI (blue), common surface antigens (CSA) on the *Salmonella* bacteria are stained in red and the *Salmonella* (pSsaG) where GFP expressing and are visible in green (GFP). This data is representative of biological replicates of different *S*. Weltevreden isolates.

All *S.* Typhimurium and *S.* Weltevreden were able to invade Hep 2 epithelial cells to some extent as monitored using fluorescent microscopy. Green intracellular *Salmonella* bacteria were observed in all cases by monitoring for the expression of GFP. *S.* Typhimurium SL1344 (pSsaG) exhibited a consistently stronger fluorescent signal at both the 2 and 6 h observation windows, compared to all *S.* Weltevreden. However, no significant difference in bacterial burden was obvious using this approach between the all *S.* Weltevreden isolates. Importantly, there were consistently lower levels of GFP-positive *S.* Weltevreden in the microscope imaging field than *S.* Typhimurium, indicating that they may be generally less invasive in this assay.

To assess the bacterial burden using an alternative and more quantitative approach, a gentamicin-killing assay, in which predominantly internalised bacteria should survive, was performed. For *S.* Typhimurium SL1344, there was a consistent increase in the number of viable internalised bacteria between 2 and 6 h post infection. At 6 h post infection, there was a statistically significant difference (SL1344: $P = 0.0001$) in the number of viable bacteria recovered compared to the 2 h time point (Figure 4.4). However no significant differences in recovered numbers were observed between 2 h and 6 h for the *S.* Weltevreden isolates. Additionally, there was a consistently lower level of invasion by all *S.* Weltevreden isolates compared to *S.* Typhimurium SL1344. Thus, these data support the observations made using microscopy.

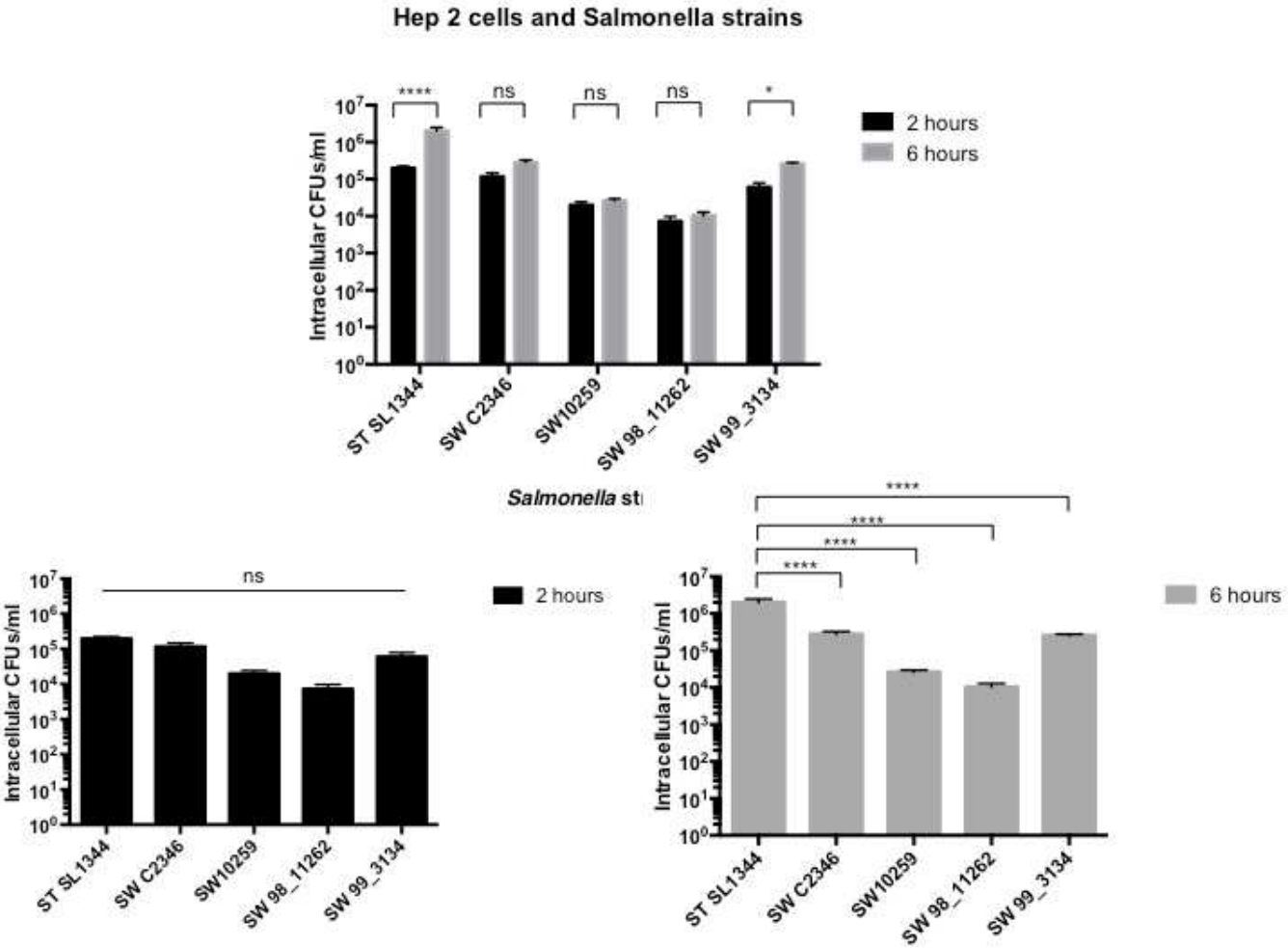
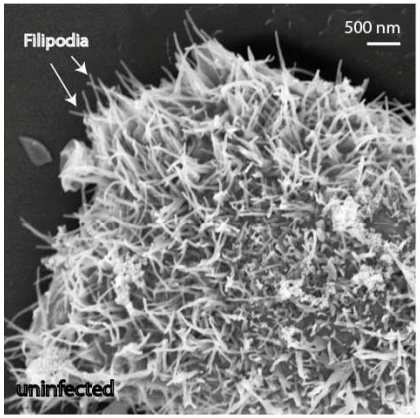

Hep 2 cells and Salmonella strains
Intracellular CFUs/ml
2 hours
6 hours
****
ns
ns
ns
*
ST SL1344
SW C2346
SW10259
SW 98_11262
SW 99_3134
Salmonella st
ns
2 hours
****
****
****
****
6 hours

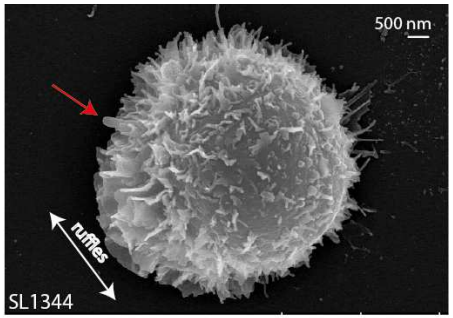
Figure 4.4: Number of viable *Salmonella* recovered in gentamicin killing assay

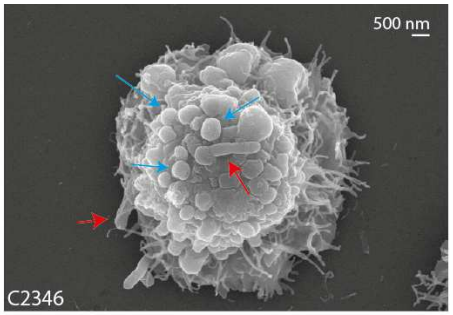
The solid bars indicate the numbers of intracellular CFU recovered 2 h (black) and 6 h (Grey) post infection from Hep 2 cells infected with *S*. Typhimurium SL1344, *S*. Weltevreden SW C2346, SW 10259 SW98_11262 or SW99_3134 (MOI, 50); the error bars indicate standard deviations. The 2-way ANOVA multiple comparisons statistics shows that *S*. Weltevreden isolates did not differ significantly between 2 h and 6 h. In contrast *S*. Typhimurium SL1344 numbers recovered increased significantly between 2 h and 6 h. *S*. Typhimurium SL1344 was also generally more invasive than the *S*. Weltevreden. Biological triplicates were used while performing this experiment.. (Symbols: ns: $p > 0.05$, *: $p \leq 0.05$, **: $P \leq 0.01$, ***: $P \leq 0.001$, ****: $P \leq 0.0001$).

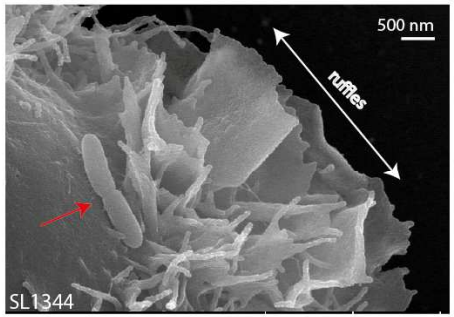
4.2.4.1 Electron microscopy

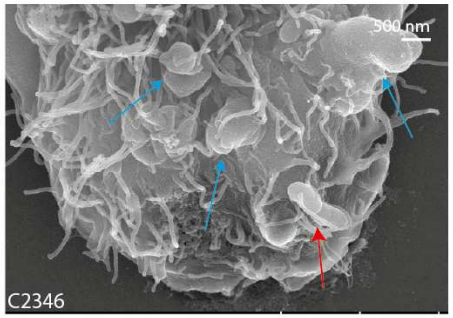
Electron microscopy was utilised to further investigate the interactions between Hep 2 cells and *S*. Weltevreden. Upon contacting host cells *S. enterica* can induce host cell membrane extensions called ruffles, which are particularly obvious upon entry into non-phagocytic cells. The appearance of such ruffles has been linked to the expression of the SPI-1 TTSS [254]. Initially *Salmonella*-Hep 2 cell interactions were investigated using Scanning Electron Microscopy (SEM). One hour post exposure of Hep 2 cells to *S*. Typhimurium SL1344 or the different *S*. Weltervreden, thin, long filopodia were observed in contact with the bacteria. These were readily visible on the Hep 2 cell surface associated with all bacteria and in the surrounding areas (Figure 4.5). These cell membrane ruffles were more obvious on the surface of *S*. Typhimurium SL1344 infected cells (Figure 4.5, middle panels). In contrast, Hep 2 cells infected with all *S*. Weltevreden isolates showed less obvious ruffling on the cell surface, particularly in contact with adherent or invading bacteria (Figure 4.5, right panels). However, no obvious differences in this phenotype were observed between the different *S*. Weltevreden isolates.

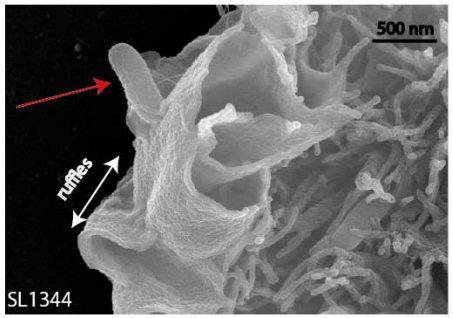

Filipodia
500 nm
uninfected


500 nm
ruffles
SL1344


500 nm
C2346


500 nm
ruffles
SL1344


500 nm
C2346


500 nm
ruffles
SL1344

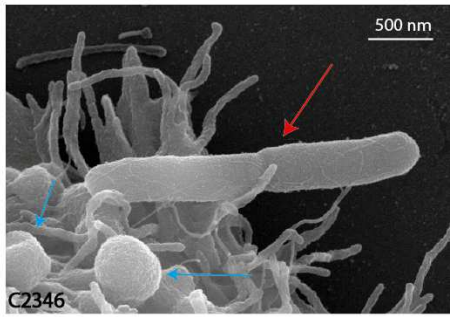

500 nm
C2346

Figure 4.5: SEM of Hep 2 cells infected with *S*. Typhimurium SL1344 or *S*. Weltevreden C2346
Red arrows show *Salmonella* and blue arrows show structures observed in *S*. Weltevreden infected cells suggestive of reduced levels of ruffling. This data is typical of replicate assays performed over 3 times with different *S*. Weltevreden isolates.

This data suggests that *S*. Weltevreden is generally less efficient at invading Hep 2 cells.

To further investigate the intracellular aspects of the infection, Hep 2 cells challenged with *S*. Typhimurum SL1344 or *S*. Weltevreden 10259 (only one representative of *S*. Weltevreden was used as no major differences between the isolates were observed in the previous experiments) were investigated using Transmission Electron Miscoscopy (TEM). Two hours post exposure to *S*. Typhimurum SL1344 and *S*. Weltevreden 10259, bacteria were routinely observed, many of which were residing within membrane-bound vacuoles (Figure 4.6). For *S*. Typhimurum SL1344 the SCV was well defined, and an enclosing membrane was clearly present. The *S*. Typhimurium had a generally healthy rod shape (Figure 4.6, panel A). Several cells with more than one bacterium within the vacuole were also observed, indicating that intracellular replication could be occurring (Figure 4.6, panel C). In contrast, cells infected with *S*. Weltevreden 10259 displayed a generally less distinct SCV, with a host membrane apparently very close to the bacterium resident within the vacuole. The bacterial cell also often exhibited an elongated form, potentially reflecting a more stressed state (Fig. 4.6, panel B). Fewer *S*. Weltervreden 10259 were observed within vacuoles or indeed within challenged Hep 2 cells.

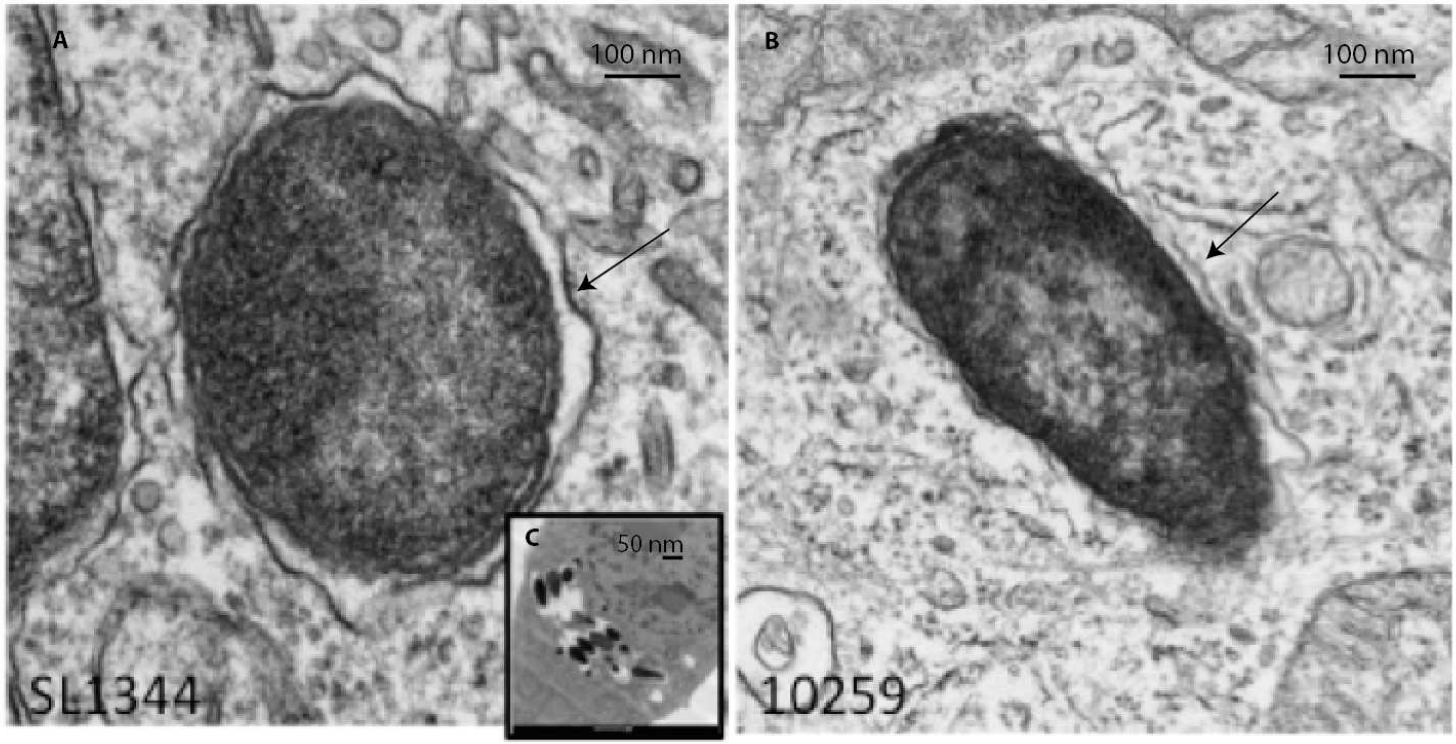


Figure 4.6: Typical TEM ultrastructures visualised within Hep 2 cells infected with *S*. Typhimurium (panel A) or *S*. Weltevreden (panel B)

A; black arrow indicating potential SCV membrane. Replication of *S*. Typhimurium Sl1344 within the SCV is represented in sub-panel C. This data is typical of replicate assays.

The observations are consistent with the concept that *S*. Weltevreden is less adapted to invasion and growth within Hep 2 cells compared to the typical *S*. Typhimurium SL1344.

4.2.5 *S*. Weltevreden in the murine model

As discussed previously there are a number of different murine models of *Salmonella* infection including the classical systemic typhoid model and the streptomycin pre-treatment model more related to gastroenteritis. Consequently, *S*. Weltervreden was evaluated in both of these murine infection models.

4.2.5.1 Systemic challenge

To determine the systemic virulence of *S*. Weltevreden, C57bl/6 (*Salmonella* susceptible, Nramp-1 negative) mice were infected intravenously with either *S*. Typhimurium SL1344, *S*. Weltevreden SW C2346 or *S*. Weltervreden SW 10259 using a dose of 2000 CFU/mice. Mice were subsequently followed over a course of 4 days to monitor clinical symptoms using an approved humane scoring method. Mice infected with the *S*. Weltevreden isolates were able to survive 4 days post infection and remained well thereafter until sacrificed. In contrast some of the C57bl/6 mice infected with *S*. Typhimurium SL1344 mice were deemed to be at the clinical endpoint in terms of severity by day 2 post infection and the others reached this state by day 4 and were sacrificed (Figure 4.7).

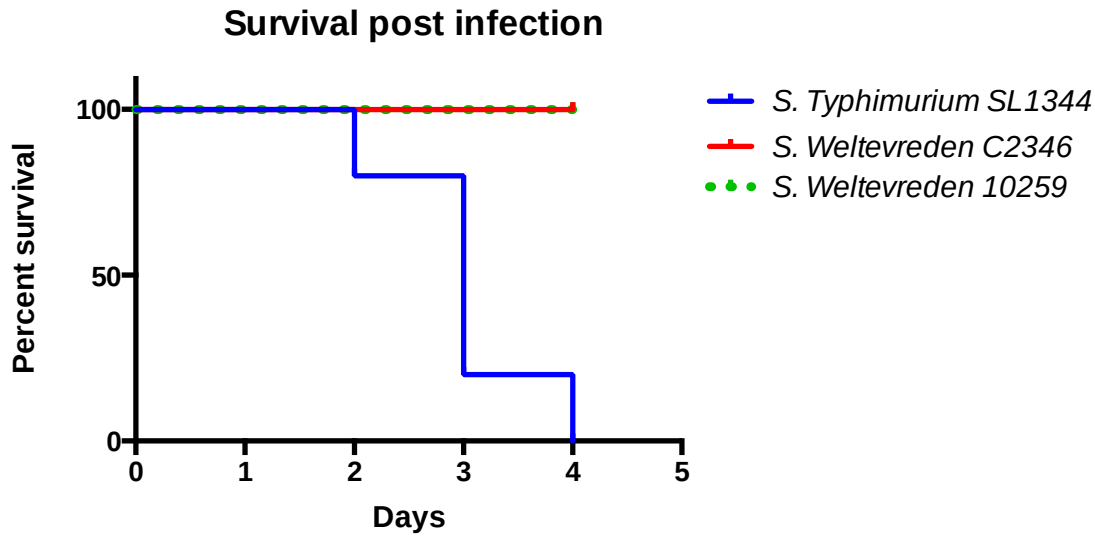


Figure 4.7: Percentage of survival of C57bl/6 mice challenged with *S*. Typhimurium SL1344, *S*. Weltevreden C2346 and *S*. Weltevreden 10259 following intravenous infection

This data is typical of replicate assays performed in triplicate.

4.2.5.2 Evaluation of *S*. Weltevreden in the streptomycin pre-treated colitis model

S. Weltevreden is an emerging cause of colitis in humans. In order to model aspects of this disease and to unravel details of the potential mechanism of infection, streptomycin pre-treated C57bl/6 mice were orally challenged to investigate the ability of *S*. Weltevreden isolates to induce an inflammatory response and cause infection in the caecum. Four days post infection, a histopathological analysis of caecum revealed pronounced inflammation characterised by oedema in the submucosa, with distinct cellular inflammatory infiltrates in the submucosa, the lamina propria, and the epithelial layer, as well as the presence of immune cells in the intestinal lumen. Crypts elongation and erosive changes in the surface epithelium were also observed. These features were present in the caecum of mice infected with either *S*. Weltevreden and *S*. Typhimurium were as the caecum of PBS challenged mice displayed no noticeable oedema or neutrophil infiltration (Figure 4.8).

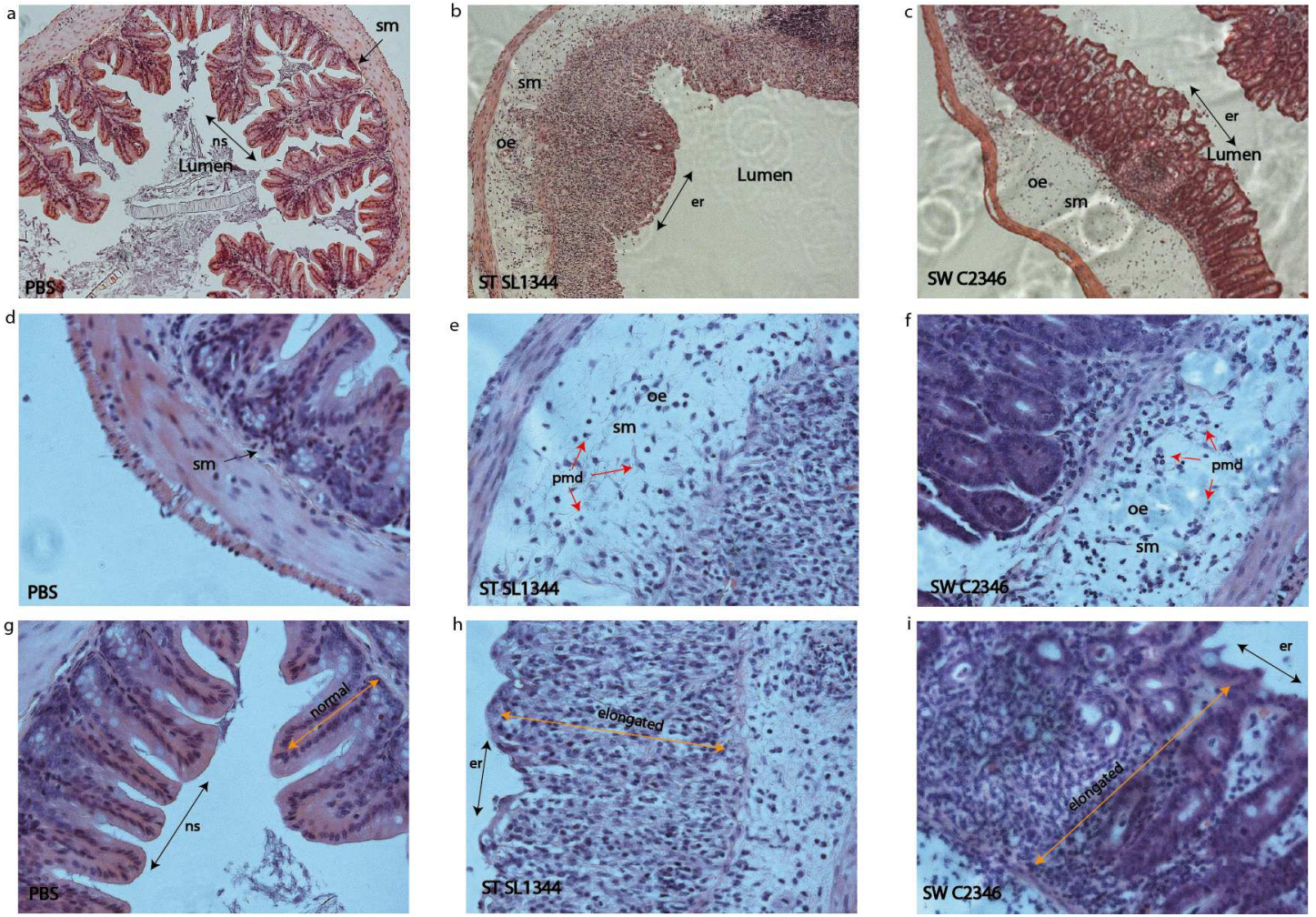

a
sm
ns
Lumen
PBS
b
sm
oe
Lumen
er
ST SL1344
c
er
Lumen
oe
sm
SW C2346
d
sm
PBS
e
oe
sm
pmd
ST SL1344
f
pmd
oe
sm
SW C2346
g
normal
ns
PBS
h
elongated
er
ST SL1344
i
er
elongated
SW C2346

Figure 4.8: Histopathological analysis of caecum sections of mice infected with *S*. Typhimurium SL1344, *S*. Weltevreden C2346 or mock-injected with PBS 4 days post infection

The top panels (a, b and c) show a representative section for each infection at magnification 10X. The mid panels (d. e and f) show details of the submucosa with infiltrates at magnification 40X. The bottom panels (g. h and i) show the structure of the crypts for each infection at magnification 40X. Abbreviations: sm: submucosa, oe: oedema, ns: normal structure, er: erosion of the membrane and pmd: polymorphonuclear leukocytes infiltrates. This data is representative of biological replicates

Thus, in contrast to the attenuated phenotype displayed by *S*. Weltevrden in the systemic murine model, similar pattern in intestinal pathology were observed between the 2 serovars in the caecum after challenge of streptomycin-treated mice.

Additionally, the levels of colonisation by the different *S. enterica* isolates was monitored by plating outweighed sections of the caecum and counting the number of surviving *Salmonella* per milligram of tissue (Figure 4.9). Plating weighed lobes of the liver also provided insights into the ability of these serovars to cause systemic infection post oral gavage in streptomycin pre-treated murine model (Figure 4.10).

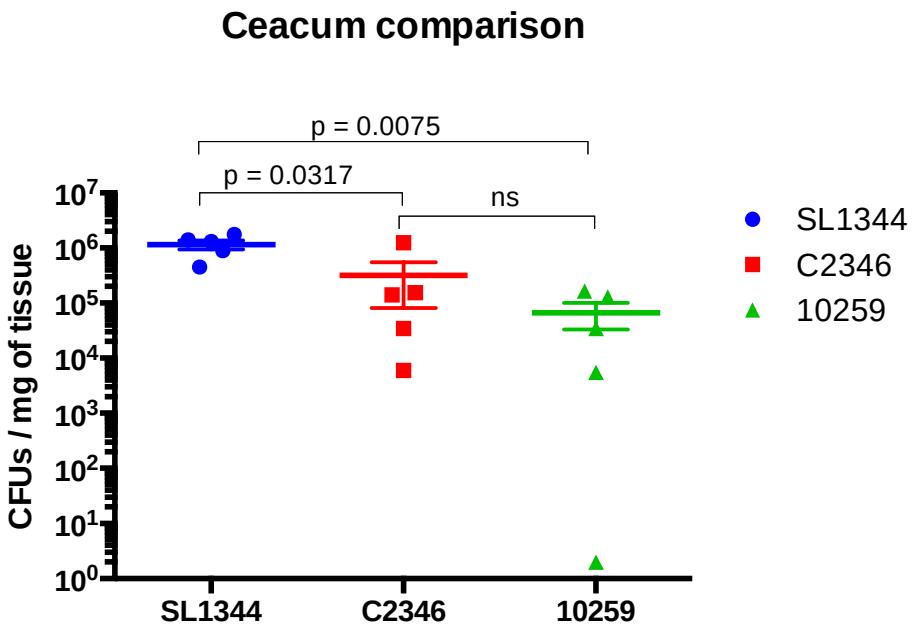


Figure 4.9: Analysis of the impact of colonisation of the ceacum by *S*. Typhimurium SL1344, *S*. Weltevreden C2346 or *S*. Weltevreden 10259 in C57bl/6 mice 4 days post infection

The Mann Whitney U T-Test shows a significant higher burden in ceaca infected with *S*. Typhimurium SL1344 compared to caecum infected with *S*. Weltevreden (p = 0.0317 and p = 0.0075). No significant difference between the different *S*. Weltevreden isolates was noted. This data is representative of biological replicates

Interestingly, *S*. Typhimurium SL1344 exhibited a consistently higher level of caecum colonisation compare to the *S*. Weltevreden isolates and these differences reached statistical significance. In contrast, there was no significant difference in caecal colonisation between the two *S*. Weltevreden isolates studied.

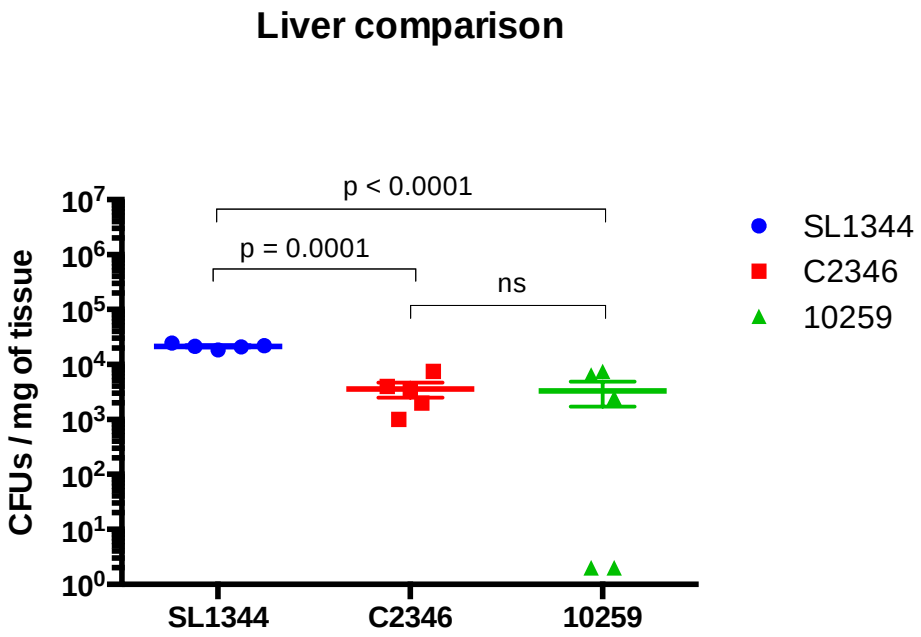


Figure 4.10: Analysis of the levels of liver colonisation of *S.* Typhimurium SL1344, *S.* Weltevreden C2346 or *S.* Weltevreden 10259 in C57bl/6 mice 4 days post infection

A significantly higher bacterial burden was detected in the livers infected with *S.* Typhimurium SL1344 compared to *S.* Weltevreden isolates as determined by the Mann Whitney U T-Test ($p = 0.0001$ and $p < 0.0001$). No significant difference was observed between the colonisation levels of the livers by the *S.* Weltevreden isolates. This data is representative of biological replicates

Thus, similarly to the data determined for colonisation in the caecum, *S.* Typhimurium SL1344 displayed a significantly higher colonisation level in murine liver compare to the *S.* Weltevreden isolates. Similarly, there was no significant difference in liver colonisation between the *S.* Weltevreden isolates.

4.3 Discussion

Phylogenetic analysis confirmed that the sequenced *S.* Weltevreden isolates fell into a monophyletic clade formed of several sub-clades. As *S.* Welterveden is an emerging cause of gastroenteritis in certain parts of the world, particularly South East Asia, it is important to link the genotype of isolates to classical phenotypic properties associated with pathogenesis. However, relatively little is known about the phenotypic properties of this serovar. Here, by comparing several *S.* Weltervreden isolates with the well-characterised *S.* Typhimurium SL1344, which was originally

isolated from a case of cattle enteritis, we were able to initiate a phenotypic characterisation of this serovar.

The initial serology confirmed the serotype of all isolates according to the Kauffman-White scheme. *S.* Weltevreden has an unusual (O10, O15) LPS antigen and r flagella type, falling into *Salmonella* Group E1. Other serovars with a similar antigenic composition include *S.* Ughelli, *S.* Elizabethville and *S.* Simi, all relatively rare in terms of their frequency of isolation. Indeed, there are virtually no reports on infections by these 3 serovars in the recent literature apart from the isolation of *S.* Simi in the Congo.

Although the *S.* Weltervreden grew well on laboratory medium and were metabolically similar to *S.* Typhimurium SL1344, they exhibited defects in their ability to interact with eukaryotic cells and in their relative virulence for mice. Indeed, the *S.* Weltevreden isolates were significantly less invasive in terms of their ability to enter Hep 2 cells compare to *S.* Typhimurium SL1344. Furthermore, membranes ruffles on the cell surface were less likely to be observed in cells infected with *S.* Weltevreden. In contrast, less robust structures were observed in the surface of the Hep 2 cells exposed to *S.* Weltervreden, usually directly associated with attached bacteria. Interestingly, *S.* Weltervreden appears to have a normal SPI-1 invasion system as far as can be deduced from simple comparative DNA analysis so it is not clear how the different invasive phenotypes are moderated genetically. *S.* Weltervreden may encode unknown effector proteins that have not yet been identified. Additionally different regulatory pathways may be in operation that impact on the expression of the SPI-1 system, although this was not investigated. Whatever, it is interesting that *S.* Weltervren does interact differently with Hep 2 cells and this could impact directly on the virulence potential of this serovar.

Similar to what was observed *in-vitro* in the Hep 2 invasion assays, the *S.* Weltevreden isolates were moderately attenuated in the mouse in both intravenous and oral streptomycin treated infection models, compare to *S.* Typhimurium SL1344. In fact, mice intravenously infected with *S.* Weltevreden were able to survive 4 days post infection while by day 2 post infection, some *S.* Typhimurium SL1344 infected mice were clinically moribund and had to be sacrificed according to the humane

protocols. Although similar levels of inflammation was observed in streptomycin treated mice infected with both serovars, there were higher systemic colony counts in the *S.* Typhimurium SL1344 infected animals. Indeed, the *S.* Weltevreden isolates displayed reduced growth and replication within the caecum and the liver of infected mice. Again, it is not clear why the *S.* Weltervreden were so attenuated in mice compared to *S.* Typhimuirum SL1344. Clearly, the reduced ability to exploit the SPI-1 invasion system could be a factor, although this system is known not to be absolutely required for mouse virulence. There was no obvious mutation, for example in a known virulence factors that could explain the attenuation, as determined from the initial interrogations of genome sequence. Clearly, this is an area worthy of further investigation in alternative systems (see next Chapter).

Some of the *S.* Weltervreden isolates under study were from cases of clinical disease and this serovar is now recognised as an emerging cause of gastroenteritis in a number of distinct geographical settings. It is interesting that even though *S.* Weltervreden shows some characteristics of attenuation on these models the isolates of this serovar are able to still cause disease in humans. Clearly, it is well established that host adaptation or even restriction is a relatively common property of *Salmonella* isolates and this may be to some degree what is being revealed by these studies. Derivatives of other *Salmonella* serovars that lack a fully functional SPI-1 or SPI-2 system can cause disease in humans. For example, *S. bongori* lacks SPI-2 but is able to cause sporadic human gastroenteritis. Thus, clearly alternative host specific mechanisms of pathogenesis occur and not all have been defined to date.

This first insight into the phenotypic characteristics of *S.* Weltevreden revealed an overall attenuated pathology compare to *S.* Typhimurium SL1344. Further studies addressing the metabolic choice of carbon sources as well as the mechanistic of epithelial cells invasion would provide a better understanding of *S.* Weltevreden interactions with the host.

5 *S.* Weltevreden in the zebrafish infection model

5.1 Introduction

In the previous chapters a phylogeny and phenotypic analysis of *S.* Weltevreden isolates was performed in an attempt to build up a database of information on this understudied serovar. *S.* Weltevreden has now been reported as a commonly isolated serovar both from the environment and from clinical cases in different parts of the world. Although the potential source of these *S.* Weltevreden isolates is varied, seafood products have been frequently implicated, implying a potential aquatic source, both fresh and salt water. These reports indicate that aquatic food sources may be an important transmission route of *S.* Weltevreden into the human population [229, 255] suggesting possibilities for marine and freshwater ecosystems as natural niches.

This link with aquatic sources indicated that *S.* Weltevreden could potentially be a coloniser of fish. Thus, it was postulated that *S.* Weltevreden could potentially infect the zebrafish, which is frequently used as an infection model [256]. Additionally, many reports of *S.* Weltevreden are specifically linked to South East Asia, a native environment of the zebrafish. The zebrafish embryo has been extensively validated as a tool for investigations into many aspects of biology ranging from development to infection and immunity. A number of different bacteria can infect zebrafish embryos and indeed adult fish [257-265], for example. *M. marinum* has been used extensively as a model both for tuberculosis and for granuloma formation studies [163, 266]. Moreover, the zebrafish has been used to explore the pathogenesis of *Salmonella* infections, in particular to analyse the early host response to infection. The availability of zebrafish lines harbouring mutations in individual genes has extended the value of such studies. A particular focus has been on the role of macrophage in the dissemination and control of *Salmonella* infection.

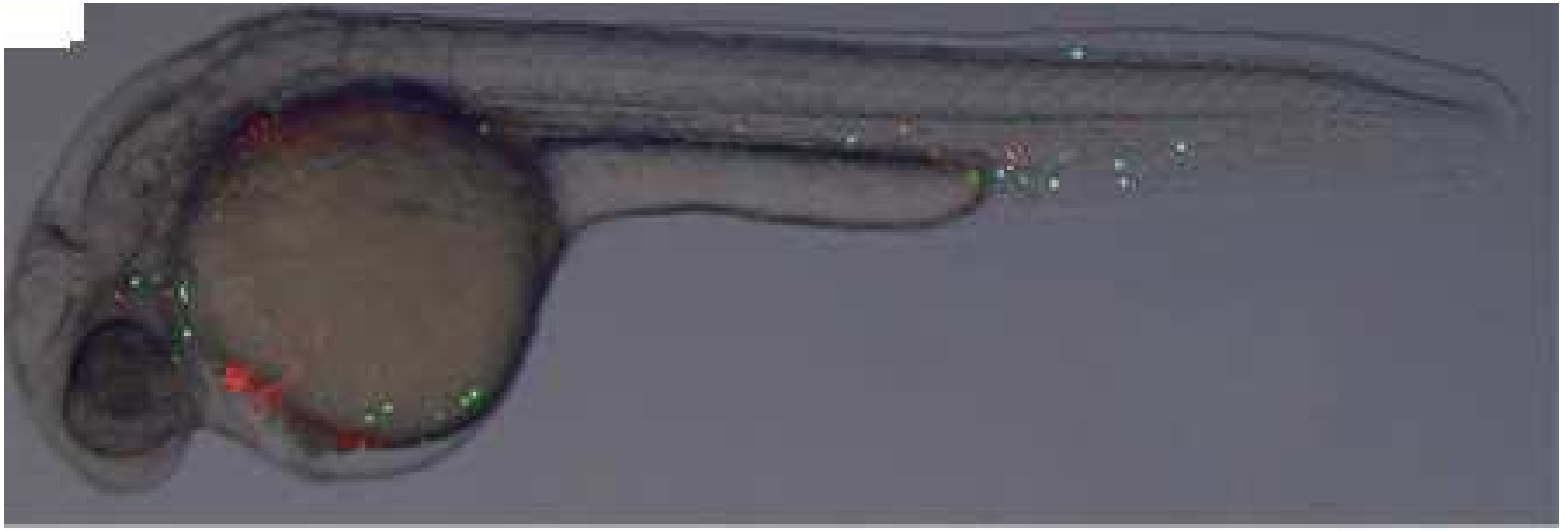


Figure 5.1: Transgenic zebrafish embryos infected with *S.* Typhimurium SL1027 48 hpi. Taken from [164]

Here, the zebrafish infection model has been utilised to investigate the ability of *S.* Weltevreden to colonise or cause significant infection in a fish species. To this end, embryos were microinjected with either *S.* Typhimurium or *S.* Weltevreden using wild type or mutant zebrafish lines and the course of infection was followed over several days.

5.2 Results

5.2.1 Bacterial growth *in-vitro* at 28°C

Since the optimal growth temperature for zebrafish is 28°C, *S.* Typhimurium SL1344, *S.* Weltevreden C2346 and 10259 were grown independently in LB broth at 28°C and their respective growth was assessed over a time course of 24 hours. The results, shown in Figure 5.2, indicated that all isolates grew at a similar growth rate in this medium. The number of bacteria per OD is similar to what observed at 37°C in the same media (see Chapter 4) but the time of replication was longer (Figure 5.2).

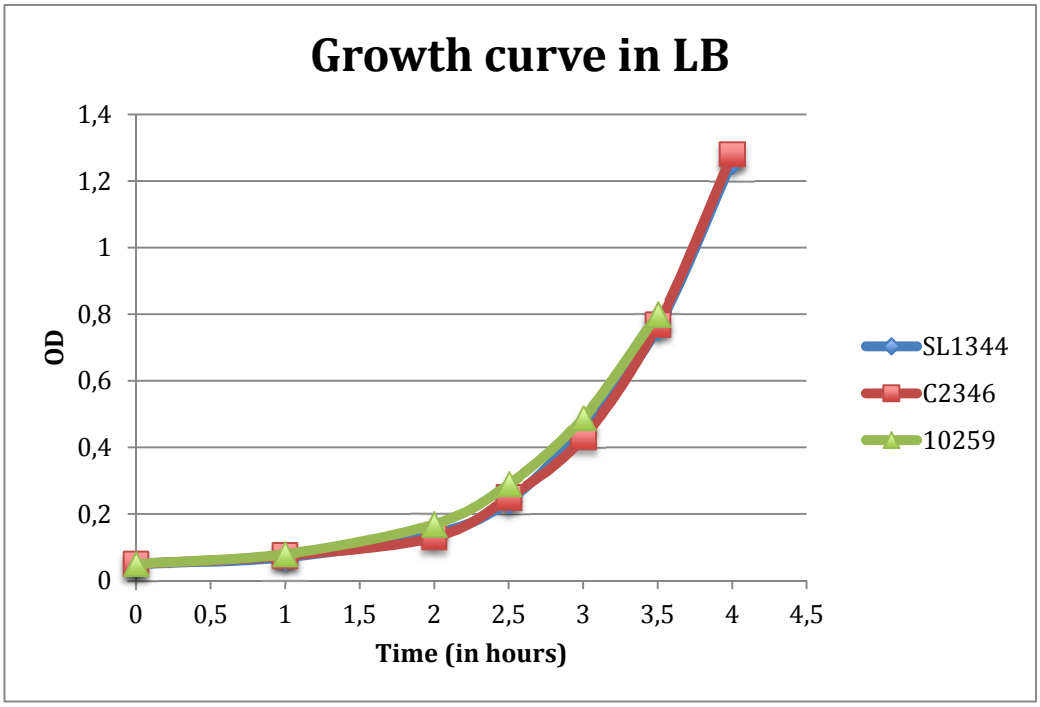


Figure 5.2: Growth curve showing the duplication time of *S.* Typhimurium SL1344 , *S.* Weltevreden C2346 and *S.* Weltevreden 10259 in LB medium at 28°C.

5.2.2 Infection challenge

To determine the ability of *S.* Weltevreden isolates to infect zebrafish, 48 hour old embryos were systemically challenged using microinjection with ~250 CFU of either *S.* Typhimurium SL1344 or *S.* Weltevreden SW C2346 or they were mock infected using microinjection alone. The embryos were subsequently monitored for up to 70 hours post infection, scoring for survival (Figure 5.3). No deaths were recorded in *S.* Weltevreden infected embryos or those mock infected. In contrast, ~50% of the embryos infected with *S.* Typhimurium were dead by 40 hours post infection (Figure 5.3).

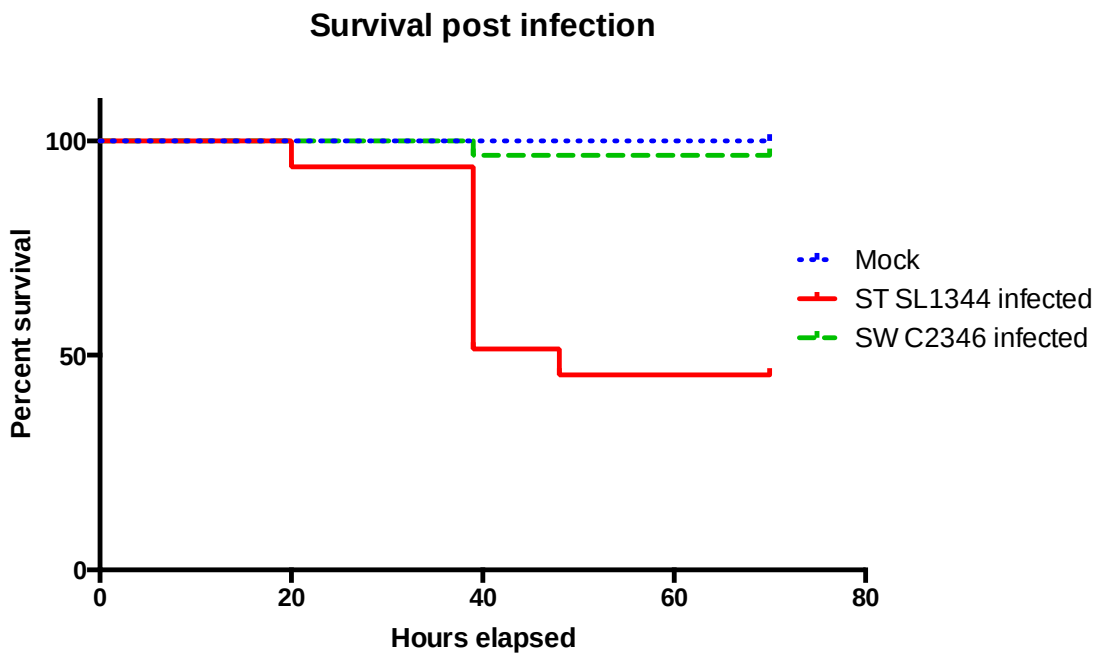


Figure 5.3: Percentage of survival of zebrafish embryos microinjected with *S.* Typhimurium SL1344 (ST), *S.* Weltevreden C2346 (SW) or mock injected

($P < 0.0001$ between *S.* Typhimurium SL1344 and *S.* Weltevreden C2346).

As *S.* Weltevreden C2346 was clearly attenuated in this model compared to *S.* Typhimurium SL1344, further experiment were performed using a ~8000 CFUs per injection dose of *S.* Weltevreden. At this dose, some of the embryos infected with *S.* Weltevreden succumbed within 24 hours post infection. However, many embryos

survived despite the use of a higher challenge dose. These data confirms a significant level of attenuation of *S.* Weltevreden C2346 in the zebrafish embryo challenge model.

5.2.3 *Salmonella* viability in the zebrafish embryo model

The numbers of *S.* Weltevreden C2346 and *S.* Typhimurium SL1344 surviving at different time points after microinjection into embryos was assessed using viable counts. This involved plating out whole embryos and counting the number of surviving *Salmonella* at different time points after challenge (Figure 5.4). The number of viable *Salmonella* recovered from the embryos was similar to the number microinjected at 6 hours post infection (data not shown in the Figure) suggesting limited growth had occurred at this time point.

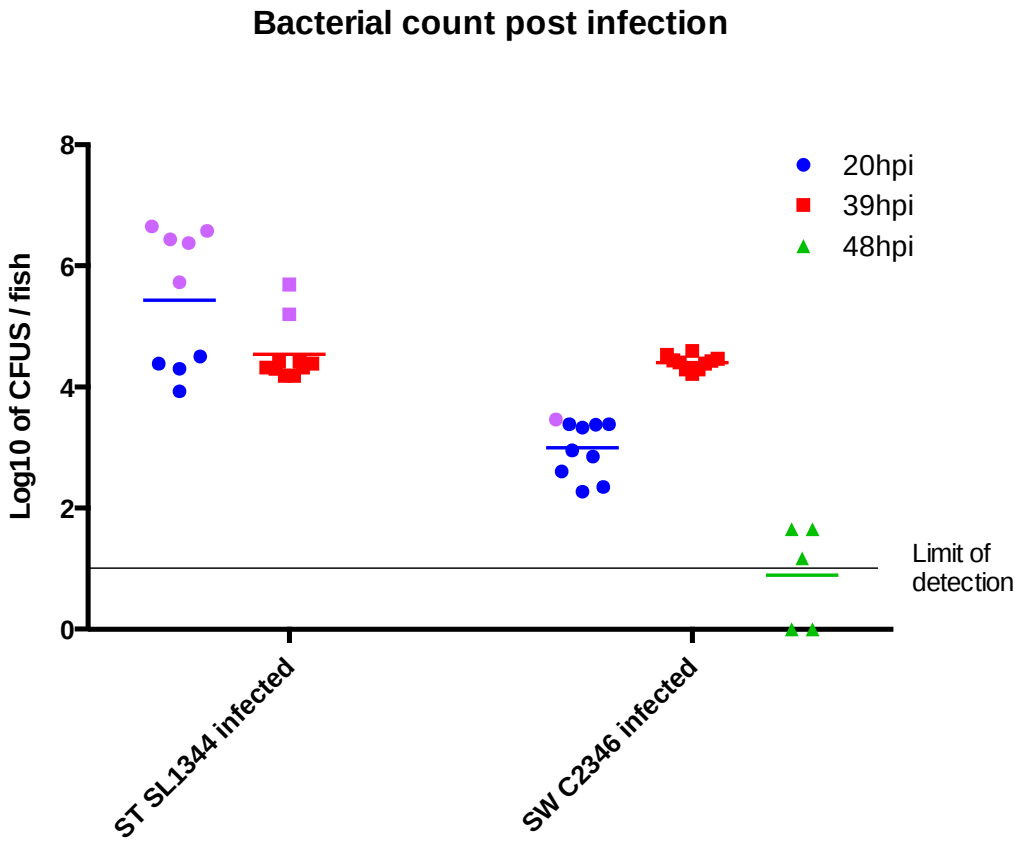


Figure 5.4: Analysis of Salmonella survival within zebrafish embryos microinjected with 250 - 300CFUs of *S.* Typhimurium SL1344 or *S.* Weltevreden C2346

Time points were 20 hpi, 39 hpi, 48 hpi. All *S*. Typhimurium challenged embryos were dead by 48 hours post challenge whereas *S*. Weltevreden C2346 was cleared at this point. The purple samples represent severely moribund fish in each group.

Because the mortality between the 2 groups was not substantially different at 20 hours post infection, statistical analysis were conducted at that time point to assess the replication of each bacterial strain within the embryos (figure 5.5). As observed earlier in the murine model, *S*. Typhimurium SL1344 displayed a significantly higher replication level in larvae compare to *S*. Weltevreden C2346.

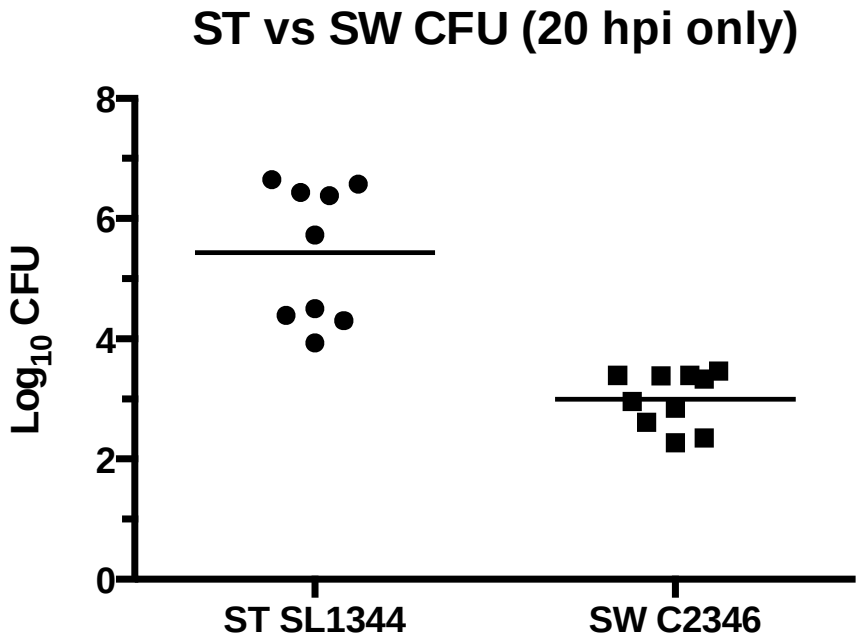


Figure 5.5: Analysis of the replication level within zebrafish embryos microinjected with 250 - 300 CFUs of *S*. Typhimurium SL 1344 (ST SL1344) and S. Weltevreden C2346 (SW C2346) 20 hours post infection - P value < 0.0001.

By 48 hours post challenge, no *S*. Typhimurium SL1344 infected embryos were left alive, whereas most *S*. Weltevreden C2346 infected embryos were viable (only 5 were plated for CFU counts). Interestingly, *S*. Weltevreden C2346 infected embryos were able to clear the *Salmonella* 48 hours post infection. Thus, although *S*. Weltevreden C2346 may be able to undergo limited replication in embryos but they

were subsequently cleared around 2 days post infection as shown in Figure 5.6 below.

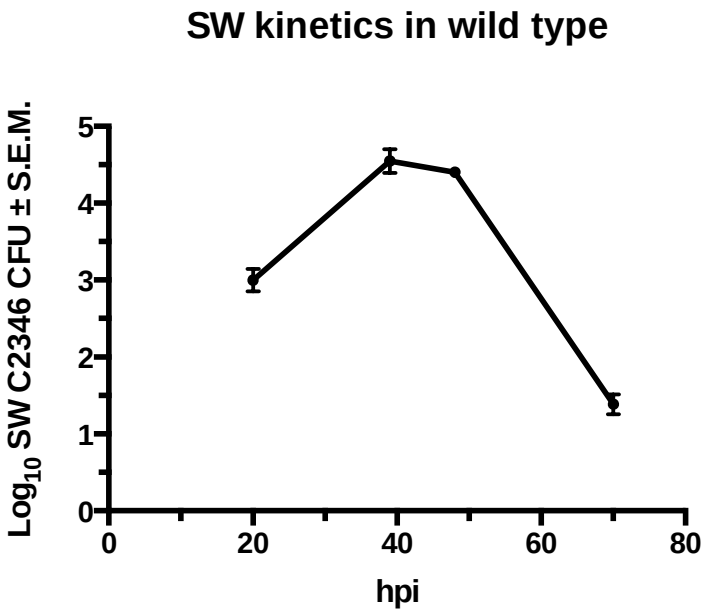


Figure 5.6: Kinetics of *S.* Weltevreden C2346 survival within zebrafish embryos over the course of the infection following challenge with 250 - 300 CFUs.

5.2.4 Survival of *S.* Weltevreden within macrophage-deficient embryos

To assess the potential role of macrophages mediated killing in the rapid clearance of *S.* Weltevreden C2346 in zebrafish embryos, we exploit an irf8 null zebrafish mutant generated through TALEN-mediated targeting. Irf 8 plays a role in myeloid cell differentiation including the differentiation of a common myeloid progenitor into a monocyte precursor cell. The nucleases targeted the region near the *irf8* translational start site, creating frame shift mutation *st95* thereby introducing premature stop

codons [225]. irf8 $^{-/-}$ zebrafish mutants are characterized by a complete lost of macrophages but an over production of neutrophils. Heterozygous mutant (irf8 $^{+/-}$) fish harbour normal levels of macrophages and neutrophils [225].

irf8 $^{+/-}$ heterozygous parents were crossed and the resulting offspring was infected with a low dose (~100 CFU) of *S.* Typhimurium SL1344, *S.* Weltevreden C2346 or mock injected. The infection was followed over a course of 70 hours and the surviving fish were genotyped. The expected distribution of genotypes in each clutch is 50% irf8 $^{+/-}$, 25% irf8 $^{-/-}$ and 25% of wild-type. Figure 5.7 summarises a compilation of the genotyping and viability data obtained in a typical experiment.

Genotypes of surviving larvae

	+/+	***st95/+***	***st95/st95***
mock	14	20	9
ST	11	19	1
SW	13	20	7

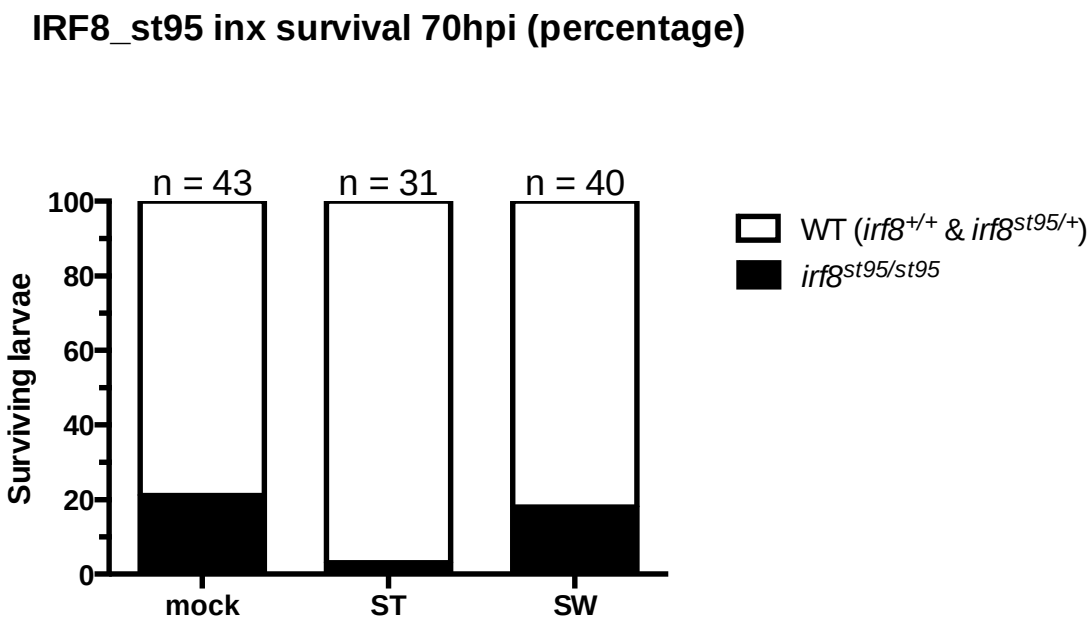


Figure 5.7:Percentage of survival of wild-type irf8+/+, heterozygous mutants irf8st95/+ and homozygous mutants irf8 st95/st95 challenged with *S.* Typhimurium SL 1344 (ST), *S.* Weltevreden C2346 (SW) or mock-injected 70 hours post infection

It is evident that irf8 $^{st95/st95}$ embryos infected with *S.* Weltevreden C2346 or mock injected have a higher survival rate than similar embryos challenged with *S.* Typhimurium SL1344. Thus, the attenuated phenotype of *S.* Weltevreden C2346 is significantly independent of the presence of macrophages. This indicates that the mechanism of attenuation is not macrophage dependent or associated.

5.3 Discussion

The initial phenotypic characterisation of *S.* Weltevreden described in the previous chapter revealed an overall attenuated phenotype in different models of disease compare to *S.* Typhimurium SL1344. *S.* Weltevreden was significantly attenuated in mouse models and in the ability to infect Hep 2 cells in an invasion assay. At this point, the mechanism of attenuation remains unknown. Interestingly, in the face of this 'attenuated' phenotype many reports are linking this serovar to serious cases of illness in humans. Thus, the correlation between these model systems and human infectivity is not absolute. Indeed, others have reported significant levels of attenuation in mouse-virulence and invasiveness with other *Salmonella,* for example the ST313 *S.* Typhimurium associated with invasive disease in sub-Saharan Africa [16, 267]. Again, in these cases the mechanisms of attenuation remain unknown, although genome degradation similar to *S.* Typhi has been reported [16]. However, we found no evidence for significant levels out of the norm for *S.* Weltevreden (see Chapter 3).

Seafood and water based products have been implicated a potential source and/or transmission route for infection with *S.* Weltevreden. This association prompted in part this investigation into the ability of *S.* Weltevreden to infect zebrafish embryos. A comparative analysis of infectivity was undertaken by setting up simultaneous infections with either *S.* Typhimurium SL1344 to *S.* Weltevreden C2346. *S.* Typhimurium SL1344 is known to be significantly virulent for zebrafish embryos and this strain and similar *S.* Typhimurium have been extensively characterised by others in this model [257].

S. Weltevreden C2346 exhibited a significant level of attenuation following microinjection into young embryos even if different doses were used. Whereas

embryos infected with *S.* Typhimurium SL1344 routinely succumbed to infection embryos similarly challenged with *S.* Weltevreden C2346 normally survived and were even able to clear the infection challenge within 48-72 hours post challenge. Importantly, this lack of virulence of *S.* Weltevreden C2346 was even present when embryos defective in macrophage production (irf8 $^{+/-}$ mutant) were challenged. This indicates that the mechanism of attenuation is significantly independent of macrophages, which are known to be key cells involved in the pathogenesis and control of *Salmonella* infections in fish and other animals [268]. The ability of macrophage-deficient (irf8 $^{-/-}$ mutants) to clear the infection suggests a potential involvement of neutrophils in controlling *S.* Weltevreden infection. Whatever, *S.* Weltevreden C2346 displays significant levels of attenuation in multiple classical virulence models yet the serovar can still cause significant disease in humans.

In a natural setting, infection may occur in adult fishes with fully developed immune systems, likely via the oral route. Thus, it would be interesting to explore the virulence of *S.* Weltevreden in adult zebrafish but this would require an animal licence not available during this study. Further studies in adult fish populations could help unravel the true relationship between *S.* Weltevreden and fish confirming whether the serovar is a natural commensal or pathogenic or simply attenuated in this particular host.

The attenuated phenotype of *S.* Weltevreden in zebrafish embryos remains of interest and further experiments are planned beyond the immediate scope and time-frame of this thesis. RNA-seq analysis will be used to explore the nature of the host response to *S.* Weltevreden compared to S. Typhimurium infection and other isolates of *S.* Weltevreden should be used in this model to assess how broadly the attenuated phenotype is present in the serovar. Additionally, other mutant zebrafish lines could be used to explore further the mechanisms of attenuation, for example to identify mutant lines that succumb to *S.* Weltevreden challenge.

6 Summary and future directions

In this study, a combination of whole genome sequencing, phylogenomics and *in-vitro*/*in-vivo* phenotyping were used to characterise the serovar *S.* Weltevreden. Additionally, a complete reference genome was generated that will prove of value for further genetic work on this serovar. This analysis revealed that the average *S.* Weltevreden genome is larger than those of many other *S. enterica* serovars, with an average size above 5,000,000 base pairs. Much of this additional DNA can be accounted for in the accessory genome, where whole prophage and additional phage-related elements are common. *S.* Elizabethville and *S.* Goldcrest were phylogenetically the closest *S. enterica* serovars to *S.* Weltevreden amongst the serovars for which other whole genome sequences are available. Interestingly, *S.* Elizabethville shares common core serological properties with *S.* Weltevreden but this serovar is not a common pathogen in humans. However, it will be interesting to see if these related serovars increase in their association with human diseases in the future.

S. Weltevreden appears to be a monophyletic serovar assignable to 2 major phylogenetic clusters of largely 'Continental cluster' and 'Island cluster' isolates. Thus, there is evidence of a significant level of geographical clustering within *S.* Weltevreden. Some geographical clustering is also detectable within the sub-phylogeny, suggesting that the *S.* Weltervreden serovar continues to evolve within a specific geographical region rather than frequently spreading from one location to another. Geographical subclustering has been detected in other serovars, including the *S.* Typhimurium ST313 clades within sub-Saharan Africa [269]. This suggests that *Salmonella* clades can become established in an environment or a population where they persist and evolve.

One hundred and twelve SNPs were found to be cluster-specific and these could be of value in epidemiological tracking. For example, it will be interesting to determine if this type of data can be used to map potential transmission routes within populations. Can solid phylogenetic links be identified between seafood and *S.*

Weltevreden in human diseases? Is there evidence of human-to-human transmission? Here, some of the cluster-associated or even private SNPs will be exploitable in simple SNP-based assays for the rapid identification of *S*. Weltevreden isolates in the field. Such approaches have been developed for other *Salmonella* serovars, including *S*. Typhi [270, 271].

It is significant that the phylogeny of *S* Weltevreden does not correlate with date of isolation, disease type or source (environment, animal, human) of the isolate, making it impossible to link particular genotypes to any disease syndrome. The inability to link genotype to human disease is intriguing and suggests that factors such as infectious dose, host susceptibility or environment could be influencing the patterns of disease. Here, more thorough epidemiological studies will be required to try to link isolates either in environment to human, animal to human or human to human transmission routes. Such studies could be performed in countries with significant levels of endemic *S*. Weltevreden disease or by analysing transmission within outbreaks, should they be identified. We are planning to perform such studies in Vietnam, where the incidence of *S*. Weltevreden is currently comparatively high. Here SNP-based assays will be applied.

The analysis of *S*. Weltevreden plasmids revealed the presence of a large and highly conserved plasmid among the isolates, despite the diversity of geographical origin. More detailed recent analysis identified a number of candidate genes that could influence the phenotype of *S*. Weltevreden and possibly persistence in the environment and host (Figure 6.1). For example, 2 large tandem non-ribosomal peptide synthetases (NRPSs) of respectively 8381 and 13304 bp were found on the plasmid adjacent to a transporter, a rare scenario in the *Salmonella* genus. Tandem NRPSs, in combination with polyketide synthases (PKs) are commonly involved in small antimicrobial peptide synthesis [272]. The latter are usually implicated in a wide range of bioactivities including antibiotic production [273, 274], toxins, immuno-suppressants [275], anti-cancer molecules [276] and anti-fungals [277, 278]. The use of small antimicrobial peptide by the bacteria remains unclear; previous studies have speculated their potential use in fighting rival microorganism [279]. Nevertheless, the fact that this plasmid is retained in the population suggests that it may offer a selective advantage to *S*. Weltevreden in some environments. The

generation if plasmid-less derivatives or even site-directed mutations in specific genes would facilitate studies on metabolism and virulence and it would be interesting to see if any plasmid-associated mutations were picked up in any future virulence screens.

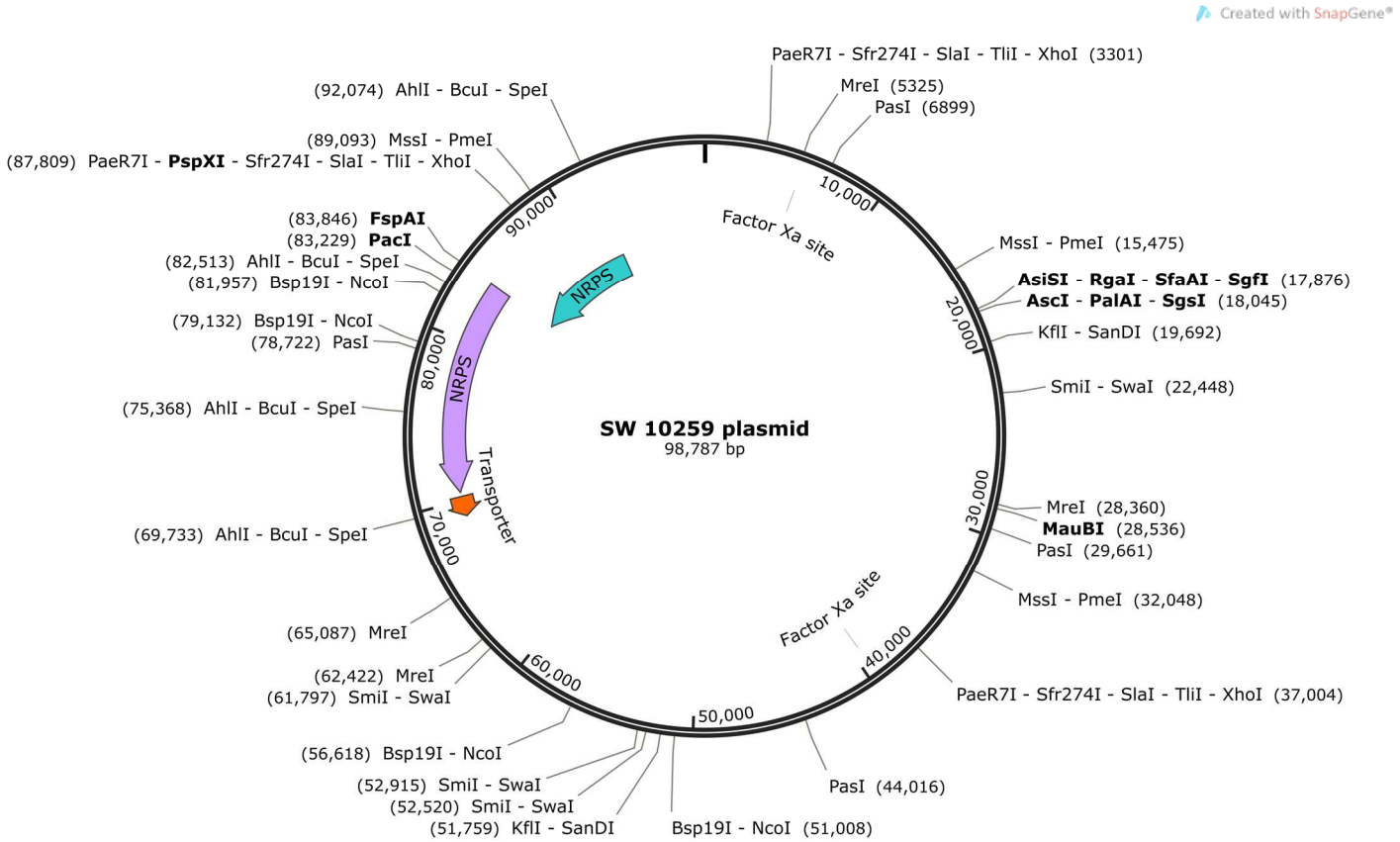


Figure 6.1: Circular map of *S*. Weltevreden 10259 plasmid

The restriction sites are marked on the plasmid. The purple and blue arrows represent the non-ribosomal peptide synthetase found on the plasmid adjacent to a putative transporter (orange arrow).

Antibiotic resistance is currently not commonly associated with *S.* Weltevreden. In fact, only a few antibiotic resistance genes were found in a few isolates. These very rare antibiotic resistant isolates harboured novel plasmids of types previously described in other bacteria. Thus, *S.* Weltevreden clearly has the capacity to acquire resistance and to evolve multiple antibiotic resistance. It will, thus, be important to maintain surveillance on the serovar in order to detect early any trend to increasing resistance.

Phenotypic characterisation of *S.* Weltevreden in comparative analysis showed an overall attenuated pathology in in different models of disease compare to *S.* Typhimurium SL1344. *S.* Weltevreden isolates were significantly less invasive in terms of their ability to enter and replicates in Hep 2 cells. Indeed, membranes ruffles, key hallmark of *Salmonella* infection on the cell surface were less likely to be observed in cells infected with *S.* Weltevreden. In contrast, less robust structures were observed in the surface of the Hep 2 cells exposed to *S.* Weltervreden. Similarly to what observed in vitro in the Hep 2 cells, *S.* Weltevreden isolates were moderately attenuated in the mouse in both intravenous and oral streptomycin treated infection models, compare to *S.* Typhimurium SL1344. In fact, mice intravenously infected with *S.* Weltevreden were able to survive 4 days post infection while by day 2 post infection, some *S.* Typhimurium SL1344 infected mice were clinically moribund.

Infection challenges of *S.* Weltevreden in zebrafish embryos also revealed an extremely attenuated phenotype and lack of virulence even in macrophage deficient zebrafish larvae. Considering that in a natural setting, infection is more likely to occur in adult fishes with fully developed immune systems via oral gavage, this exceptionally attenuated phenotype in laboratory setting suggests that the serovar is likely to be non-pathogenic for adult zebrafish. Unfortunately, we were unable to assess the virulence of *S.* Weltevreden in adult fish due to constraints on our animal licence but this work will be progressed in the future. Further studies in fish will be required to explain the prevalence of *S.* Weltevreden in marine products. In addition to the lack of virulence observed with this serovar in the fish, the ability of macrophage-deficient mutants to clear the infection suggest a critical involvement of neutrophils in controlling *S.* Weltevreden infection.

Despite the lack of antimicrobial resistance reported and the high level of attenuation reported in all diseased models explored in this study, more reports are linking *S.* Weltevreden to serious cases of foodborne illness and its predominance as a foodborne pathogen, particularly in the South-East Asian region. These data support the case for additional studies must be undertaken on the pathogenicity of *S.* Weltevreden. However, the options for studies on human disease are limited. One option might be to exploit human models based on differentiated human stem cell. New advances have facilitated the generation of different cell types and organoids from human induced pluripotent stem cells [280-284]. These include human macrophages and intestinal organoids that have both been used previously to explore *Salmonella* pathogenicity [285-287]. It would be interesting to evaluate the interaction of *S.* Weltevreden in such systems and studies of this type are planned. Alternatively human challenge studies similar to those undertaken with *S.* Typhi could be performed with *S.* Weltevreden [288].

A number of other questions remain to be answered including:

Does the accessory genome impact on *S.* Weltevreden infection and regional spread? Despite the low level of antibiotic resistance reported in *S.* Weltevreden, the serovar is successful in colonising, causing disease and spreading, as observed throughout South-East Asia. It is not clear yet how this has been driven but this is unlikely to be only associated with the core genome, which is broadly shared across *S. enterica*. Phage and other mobile elements are known to be key drivers of diversity and evolution within *S. enterica* [128] and such elements might be worth interrogating experimentally in order to get a better answer on the geographical predominance of this serovar. However, additional sequencing coupled with a larger sample collection will be required to better characterise the accessory genome. The work presented here might be the foundation for further functional genomic work, including mutagenesis, RNA-seq and proteomics aiming to link the genotype to different phenotypes.

Does *S.* Weltevreden exploit alternative regulatory pathway to invade and cause disease? *S* Weltevreden appears to have normal SPI-1 locus as determined by DNA sequencing and comparative DNA analysis. However, defects in invasion and/or

intracellular colonisation were observed in both *in-vitro* and *in-vivo* systems. In addition, the utilisation of alternative carbon sources commonly associated with soil and plant organism might suggest that *S.* Weltervreden inhabits different host environments compared to other *Salmonella*. For example, the accessory genome may encode novel effector proteins that could either exploit alternative survival pathways. Alternatively *S.* Weltevreden might have evolved novel regulatory systems that impact on the expression of the SPI-1 and other virulence-associated system. Experiments manipulating different pathways alongside RNA-seq analysis might begin to unravel the mechanisms of *S.* Weltevreden interaction with the host.

Are neutrophils key in controlling *S.* Weltevreden infection? In our zebrafish infection studies, *S.* Weltevreden control appeared to coincide with the peak of the emergency granulopoiesis, previously reported after Salmonella infection in larvae by others [289]. We found that macrophage-deficient zebrafish larvae survived *S.* Weltevreden infection, suggesting that neutrophils might be essential in controlling such infections. Future experiments addressing the role of neutrophils in *S.* Weltevreden infection might include using mutations that block neutrophil development in zebrafish embryos, such as Csf3r [290] or Runx1 [291]. Here, morpholinos could also be used to inhibit neutrophil effector mechanisms like reactive oxidase production [291]. Other approaches could involve depleting neutrophils in mice with antibody.

Is *S.* Weltevreden a commensal in zebrafish and potentially other marine animals? Initial experiments presented here on *S.* Weltevreden infections in zebrafish embryos suggest an extremely attenuated phenotype. Monitored experiment in adult population replication the natural scenario in the wild could provide insight in the relationship between *S.* Weltevreden and marine animals. Indeed, it is possible that *S.* Weltevreden is a commensal in such populations, able to colonise without causing disease. This would potentiate any threat to humans who consumed contaminated food. Thus, it might be worth performing extensive environmental studies, particularly in potentially contaminated environments, to try to capture the true habitat of *S.* Weltevreden. Finally, it is clear that *S.* Weltevreden is emerging as a potential threat to human health in many parts of the world and that we know very little about the epidemiology of disease and the pathogenicity of the

serovar. Clearly there is a need for continuing studies in this serovar, some as outlined here.

7 References

1. World Health Organization, *WHO: Foodborne Zoonoses.* 2015.

2. Baize, S., et al., *Emergence of Zaire Ebola virus disease in Guinea.* N Engl J Med, 2014. **371**(15): p. 1418-25.

3. Maganga, G.D., et al., *Ebola virus disease in the Democratic Republic of Congo.* N Engl J Med, 2014. **371**(22): p. 2083-91.

4. Camacho, A., et al., *Potential for large outbreaks of Ebola virus disease.* Epidemics, 2014. **9**: p. 70-8.

5. Taylor, M.E. and B.A. Oppenheim, *Hospital-acquired infection in elderly patients.* J Hosp Infect, 1998. **38**(4): p. 245-60.

6. Reed, D. and S.A. Kemmerly, *Infection control and prevention: a review of hospital-acquired infections and the economic implications.* Ochsner J, 2009. **9**(1): p. 27-31.

7. Levy, S.B. and B. Marshall, *Antibacterial resistance worldwide: causes, challenges and responses.* Nat Med, 2004. **10**(12 Suppl): p. S122-9.

8. Kidd, B.A., et al., *Unifying immunology with informatics and multiscale biology.* Nat Immunol, 2014. **15**(2): p. 118-27.

9. Raskin, D.M., et al., *Bacterial genomics and pathogen evolution.* Cell, 2006. **124**(4): p. 703-14.

10. Desai, P.T., et al., *Evolutionary Genomics of Salmonella enterica Subspecies.* MBio, 2013. **4**(2).

11. Lawley, T.D., et al., *Genome-wide screen for Salmonella genes required for long-term systemic infection of the mouse.* PLoS Pathog, 2006. **2**(2): p. e11.

12. Fabrega, A. and J. Vila, *Salmonella enterica serovar Typhimurium skills to succeed in the host: virulence and regulation.* Clin Microbiol Rev, 2013. **26**(2): p. 308-41.

13. Everitt, A.R., et al., *IFITM3 restricts the morbidity and mortality associated with influenza.* Nature, 2012. **484**(7395): p. 519-23.

14. Pham, T.A., et al., *Epithelial IL-22RA1-mediated fucosylation promotes intestinal colonization resistance to an opportunistic pathogen.* Cell Host Microbe, 2014. **16**(4): p. 504-16.

15. Clare, S., et al., *Enhanced susceptibility to Citrobacter rodentium infection in microRNA-155-deficient mice.* Infect Immun, 2013. **81**(3): p. 723-32.

16. Okoro, C.K., et al., *Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa.* Nat Genet, 2012. **44**(11): p. 1215-21.

17. He, M., et al., *Emergence and global spread of epidemic healthcare-associated Clostridium difficile.* Nat Genet, 2013. **45**(1): p. 109-13.

18. Holt, K.E., et al., *Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi.* BMC Genomics, 2009. **10**: p. 36.

19. World Health Organization, *WHO: Diarrhoeal Disease.* 2015.

20. World Health Organization, *Global Water Supply and Sanitation Assessment.* 2000.

21. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* Lancet, 2013. **382**(9888): p. 209-22.

22. Reddy, E.A., A.V. Shaw, and J.A. Crump, *Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis.* Lancet Infect Dis, 2010. **10**(6): p. 417-32.

23. Nygård, K., et al., *Outbreak of Salmonella Thompson infections linked to imported rucola lettuce.* Foodborne Pathogens and Disease, 2008. **5**(2): p. 165-173.

24. Dodson, K. and J. LeJeune, *Escherichia coli O157: H7, Campylobacter jejuni, and Salmonella prevalence in cull dairy cows marketed in northeastern Ohio.* Journal of Food Protection®, 2005. **68**(5): p. 927-931.

25. Kotloff, K.L., et al., *Global burden of Shigella infections: implications for vaccine development and implementation of control strategies.* Bull World Health Organ, 1999. **77**(8): p. 651-66.

26. Ranjbar, R., et al., *Increased isolation and characterization of Shigella sonnei obtained from hospitalized children in Tehran, Iran.* J Health Popul Nutr, 2008. **26**(4): p. 426-30.

27. Allos, B.M., *Campylobacter jejuni Infections: update on emerging issues and trends.* Clin Infect Dis, 2001. **32**(8): p. 1201-6.

28. McFarland, L.V., *Epidemiology, risk factors and treatments for antibiotic-associated diarrhea.* Dig Dis, 1998. **16**(5): p. 292-307.

29. Viswanathan, V.K., M.J. Mallozzi, and G. Vedantam, *Clostridium difficile infection: An overview of the disease and its pathogenesis, epidemiology and interventions.* Gut Microbes, 2010. **1**(4): p. 234-242.

30. Warny, M., et al., *Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe.* Lancet, 2005. **366**(9491): p. 1079-84.

31. Loo, V.G., et al., *A predominantly clonal multi-institutional outbreak of Clostridium difficile-associated diarrhea with high morbidity and mortality.* N Engl J Med, 2005. **353**(23): p. 2442-9.

32. Mulvey, M.R., et al., *Hypervirulent Clostridium difficile strains in hospitalized patients, Canada.* Emerg Infect Dis, 2010. **16**(4): p. 678-81.

33. Pothoulakis, C., *Effects of Clostridium difficile toxins on epithelial cell barrier.* Ann N Y Acad Sci, 2000. **915**: p. 347-56.

34. Nastasi, A., C. Mammina, and L. Salsa, *Outbreak of Salmonella enteritis bongori 48:z35:- in Sicily.* Euro Surveill, 1999. **4**(9): p. 97-98.

35. Giammanco, G.M., et al., *Persistent endemicity of Salmonella bongori 48:z(35):--in Southern Italy: molecular characterization of human, animal, and environmental isolates.* J Clin Microbiol, 2002. **40**(9): p. 3502-5.

36. Doolittle, R.F., et al., *Determining divergence times of the major kingdoms of living organisms with a protein clock.* Science, 1996. **271**(5248): p. 470-7.

37. Klemm, E.J., Wong, V. K. and Dougan, G, *Salmonella Genomes in the Context of Lifestyle.* . eLS, 2015: p. 1-9.

38. Brenner, F.W., et al., *Salmonella nomenclature.* J Clin Microbiol, 2000. **38**(7): p. 2465-7.

39. Popoff, M.Y., J. Bockemuhl, and L.L. Gheesling, *Supplement 2002 (no. 46) to the Kauffmann-White scheme.* Res Microbiol, 2004. **155**(7): p. 568-70.

40. Szu, S.C. and S. Bystricky, *Physical, chemical, antigenic, and immunologic characterization of polygalacturonan, its derivatives, and Vi antigen from Salmonella typhi.* Methods Enzymol, 2003. **363**: p. 552-67.

41. Szu, S.C., et al., *Relation between structure and immunologic properties of the Vi capsular polysaccharide.* Infect Immun, 1991. **59**(12): p. 4555-61.

42. Edsall, G., et al., *Studies on infection and immunity in experimental typhoid fever. I. Typhoid fever in chimpanzees orally infected with Salmonella typhosa.* J Exp Med, 1960. **112**: p. 143-66.

43. Santhanam, S.K., et al., *The virulence polysaccharide Vi released by Salmonella Typhi targets membrane prohibitin to inhibit T-cell activation.* J Infect Dis, 2014. **210**(1): p. 79-88.

44. Bhan, M.K., et al., *Association between Helicobacter pylori infection and increased risk of typhoid fever.* J Infect Dis, 2002. **186**(12): p. 1857-60.

45. Dunstan, S.J., et al., *Genes of the class II and class III major histocompatibility complex are associated with typhoid fever in Vietnam.* J Infect Dis, 2001. **183**(2): p. 261-268.

46. Dunstan, S.J., et al., *Variation at HLA-DRB1 is associated with resistance to enteric fever.* Nat Genet, 2014. **46**(12): p. 1333-6.

47. Dharmana, E., et al., *HLA-DRB1*12 is associated with protection against complicated typhoid fever, independent of tumour necrosis factor alpha.* Eur J Immunogenet, 2002. **29**(4): p. 297-300.

48. Parry, C.M., et al., *Typhoid fever.* N Engl J Med, 2002. **347**(22): p. 1770-82.

49. World Health Organization, *Background Document: The diagnosis, treatment and prevention of typhoid fever.* 2003.

50. Bhan, M.K., R. Bahl, and S. Bhatnagar, *Typhoid and paratyphoid fever.* Lancet, 2005. **366**(9487): p. 749-62.

51. Lang, R., et al., *Salmonella paratyphi C osteomyelitis: report of two separate episodes 17 years apart.* Scand J Infect Dis, 1992. **24**(6): p. 793-6.

52. Lanata, C.F., et al., *Vi serology in detection of chronic Salmonella typhi carriers in an endemic area.* Lancet, 1983. **2**(8347): p. 441-3.

53. Sirinavin, S., L. Pokawattana, and A. Bangtrakulnondh, *Duration of nontyphoidal Salmonella carriage in asymptomatic adults.* Clin Infect Dis, 2004. **38**(11): p. 1644-5.

54. Gordon, M.A. and S.M. Graham, *Invasive salmonellosis in Malawi.* J Infect Dev Ctries, 2008. **2**(6): p. 438-42.

55. Kariuki, S., et al., *Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission?* J Med Microbiol, 2006. **55**(Pt 5): p. 585-91.

56. Dhanoa, A. and Q.K. Fatt, *Non-typhoidal Salmonella bacteraemia: epidemiology, clinical characteristics and its' association with severe immunosuppression.* Ann Clin Microbiol Antimicrob, 2009. **8**: p. 15.

57. Rabsch, W., et al., *Salmonella enterica serotype Typhimurium and its host-adapted variants.* Infect Immun, 2002. **70**(5): p. 2249-55.

58. Achtman, M., et al., *Population structures in the SARA and SARB reference collections of Salmonella enterica according to MLST, MLEE and microarray hybridization.* Infect Genet Evol, 2013. **16**: p. 314-25.

59. Achtman, M., et al., *Multilocus sequence typing as a replacement for serotyping in Salmonella enterica.* PLoS Pathog, 2012. **8**(6): p. e1002776.

60. Mayer, L.W., *Use of plasmid profiles in epidemiologic surveillance of disease outbreaks and in tracing the transmission of antibiotic resistance.* Clin Microbiol Rev, 1988. **1**(2): p. 228-43.

61. Hielm, S., et al., *Genomic analysis of Clostridium botulinum group II by pulsed-field gel electrophoresis.* Appl Environ Microbiol, 1998. **64**(2): p. 703-8.

62. Liu, S.L., et al., *Bacterial phylogenetic clusters revealed by genome structure.* J Bacteriol, 1999. **181**(21): p. 6747-55.

63. Cooke, F.J., et al., *Characterization of the genomes of a diverse collection of Salmonella enterica serovar Typhimurium definitive phage type 104.* J Bacteriol, 2008. **190**(24): p. 8155-62.

64. Soler-Garcia, A.A., et al., *Differentiation of Salmonella strains from the SARA, SARB and SARC reference collections by using three genes PCR-RFLP and the 2100 Agilent Bioanalyzer.* Front Microbiol, 2014. **5**: p. 417.

65. Hughes, L.A., et al., *Multi-locus sequence typing of Salmonella enterica serovar Typhimurium isolates from wild birds in northern England suggests host-adapted strain.* Lett Appl Microbiol, 2010. **51**(4): p. 477-9.

66. Francisco, A.P., et al., *Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach.* BMC Bioinformatics, 2009. **10**: p. 152.

67. Feil, E.J., et al., *eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data.* J Bacteriol, 2004. **186**(5): p. 1518-30.

68. McQuiston, J., et al., *Sequencing and comparative analysis of flagellin genes fliC, fljB, and flpA from Salmonella.* Journal of clinical microbiology, 2004. **42**(5): p. 1923-1932.

69. Marks, L.R., R.M. Reddinger, and A.P. Hakansson, *High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in Streptococcus pneumoniae.* MBio, 2012. **3**(5).

70. Seth-Smith, H.M., et al., *Structure, diversity, and mobility of the Salmonella pathogenicity island 7 family of integrative and conjugative elements within Enterobacteriaceae.* Journal of bacteriology, 2012. **194**(6): p. 1494-1504.

71. Fookes, M., et al., *Salmonella bongori provides insights into the evolution of the Salmonellae.* PLoS Pathog, 2011. **7**(8): p. e1002191.

72. Petty, N.K., et al., *The Citrobacter rodentium genome sequence reveals convergent evolution with human pathogenic Escherichia coli.* Journal of Bacteriology, 2010. **192**(2): p. 525-538.

73. Lesic, B., et al., *Quorum sensing differentially regulates Pseudomonas aeruginosa type VI secretion locus I and homologous loci II and III, which are required for pathogenesis.* Microbiology, 2009. **155**(9): p. 2845-2855.

74. Hansen-Wester, I., D. Chakravortty, and M. Hensel, *Functional transfer of Salmonella pathogenicity island 2 to Salmonella bongori and Escherichia coli.* Infection and immunity, 2004. **72**(5): p. 2879-2888.

75. Lawrence, J.G. and J.R. Roth, *Evolution of coenzyme B12 synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex.* Genetics, 1996. **142**(1): p. 11-24.

76. Porwollik, S., R.M.-Y. Wong, and M. McClelland, *Evolutionary genomics of Salmonella: gene acquisitions revealed by microarray analysis.* Proceedings of the National Academy of Sciences, 2002. **99**(13): p. 8956-8961.

77. Kauffmann, F., *[Differential diagnosis and pathogenicity of Samonella java and Salmonella paratyphi B]*. Zeitschrift für Hygiene und Infektionskrankheiten, 1955. **141**(6): p. 546-550.

78. Chen, Y.-T., et al., *Genomic diversity of citrate fermentation in Klebsiella pneumoniae.* BMC microbiology, 2009. **9**(1): p. 168.

79. Eswarappa, S.M., et al., *lac repressor is an antivirulence factor of Salmonella enterica: its role in the evolution of virulence in Salmonella.* PLoS One, 2009. **4**(6): p. e5789.

80. Chaudhuri, R.R., et al., *Comprehensive assignment of roles for Salmonella typhimurium genes in intestinal colonization of food-producing animals.* PLoS Genet, 2013. **9**(4): p. e1003456.

81. Mogensen, T.H., *Pathogen recognition and inflammatory signaling in innate immune defenses.* Clinical microbiology reviews, 2009. **22**(2): p. 240-273.

82. Takaya, A., et al., *Derepression of Salmonella pathogenicity island 1 genes within macrophages leads to rapid apoptosis via caspase□1□and caspase□3□dependent pathways.* Cellular microbiology, 2005. **7**(1): p. 79-90.

83. Kage, H., et al., *Coordinated regulation of expression of Salmonella pathogenicity island 1 and flagellar type III secretion systems by ATP-dependent ClpXP protease.* Journal of bacteriology, 2008. **190**(7): p. 2470-2478.

84. Hansen-Wester, I. and M. Hensel, *Salmonella pathogenicity islands encoding type III secretion systems.* Microbes and Infection, 2001. **3**(7): p. 549-559.

85. Zhou, D., et al., *A Salmonella inositol polyphosphatase acts in conjunction with other bacterial effectors to promote host cell actin cytoskeleton rearrangements and bacterial internalization.* Molecular microbiology, 2001. **39**(2): p. 248-260.

86. Patel, J.C. and J.E. Galán, *Differential activation and function of Rho GTPases during Salmonella–host cell interactions.* The Journal of cell biology, 2006. **175**(3): p. 453-463.

87. Malik-Kale, P., et al., *Salmonella–at home in the host cell.* Salmonella host-pathogen interactions, 2011: p. 31.

88. Cooper, K.G., et al., *Activation of Akt by the bacterial inositol phosphatase, SopB, is wortmannin insensitive.* PloS one, 2011. **6**(7): p. e22260.

89. Lawhon, S.D., et al., *Role of SPI-1 secreted effectors in acute bovine response to Salmonella enterica Serovar Typhimurium: a systems biology analysis approach.* PLoS One, 2011. **6**(11): p. e26869.

90. Hernandez, L.D., et al., *Salmonella modulates vesicular traffic by altering phosphoinositide metabolism.* Science, 2004. **304**(5678): p. 1805-1807.

91. Collier-Hyams, L.S., et al., *Cutting edge: Salmonella AvrA effector inhibits the key proinflammatory, anti-apoptotic NF-κB pathway.* The Journal of Immunology, 2002. **169**(6): p. 2846-2850.

92. Ye, Z., et al., *Salmonella effector AvrA regulation of colonic epithelial cell inflammation by deubiquitination.* The American journal of pathology, 2007. **171**(3): p. 882-892.

93. Sun, J., et al., *Bacterial activation of beta-catenin signaling in human epithelia.* Am J Physiol Gastrointest Liver Physiol, 2004. **287**(1): p. G220-7.

94. Hicks, S.W. and J.E. Galán, *Hijacking the host ubiquitin pathway: structural strategies of bacterial E3 ubiquitin ligases.* Current opinion in microbiology, 2010. **13**(1): p. 41-46.

95. La Ragione, R., W. Cooley, and M.J. Woodward, *The role of fimbriae and flagella in the adherence of avian strains of Escherichia coli O78: K80 to tissue culture cells and tracheal and gut explants.* Journal of Medical Microbiology, 2000. **49**(4): p. 327-338.

96. Lasaro, M.A., et al., *F1C fimbriae play an important role in biofilm formation and intestinal colonization by the Escherichia coli commensal strain Nissle 1917.* Applied and environmental microbiology, 2009. **75**(1): p. 246-251.

97. Shah, D.H., et al., *Cell invasion of poultry-associated Salmonella enterica serovar Enteritidis isolates is associated with pathogenicity, motility and proteins secreted by the type III secretion system.* Microbiology, 2011. **157**(5): p. 1428-1445.

98. Ren, T., et al., *Flagellin-deficient Legionella mutants evade caspase-1-and Naip5-mediated macrophage immunity.* 2006.

99. Sun, Y.-H., H.G. Rolán, and R.M. Tsolis, *Injection of flagellin into the host cell cytosol by Salmonella enterica serotype Typhimurium.* Journal of Biological Chemistry, 2007. **282**(47): p. 33897-33901.

100. e Sousa, C.R., *Dendritic cells in a mature age.* Nature Reviews Immunology, 2006. **6**(6): p. 476-483.

101. Fields, P.I., et al., *Mutants of Salmonella typhimurium that cannot survive within the macrophage are avirulent.* Proc Natl Acad Sci U S A, 1986. **83**(14): p. 5189-93.

102. Gerlach, R.G., et al., *Salmonella Pathogenicity Island 4 encodes a giant non□fimbrial adhesin and the cognate type 1 secretion system.* Cellular microbiology, 2007. **9**(7): p. 1834-1850.

103. Kingsley, R.A., et al., *Molecular and phenotypic analysis of the CS54 island of Salmonella enterica serotype typhimurium: identification of intestinal colonization and persistence determinants.* Infect Immun, 2003. **71**(2): p. 629-40.

104. Müller, P., D. Chikkaballi, and M. Hensel, *Functional dissection of SseF, a membrane-integral effector protein of intracellular Salmonella enterica.* PloS one, 2012. **7**(4): p. e35004-e35004.

105. Abrahams, G.L., P. Müller, and M. Hensel, *Functional Dissection of SseF, a Type III Effector Protein Involved in Positioning the Salmonella□Containing Vacuole.* Traffic, 2006. **7**(8): p. 950-965.

106. Deiwick, J., et al., *The translocated Salmonella effector proteins SseF and SseG interact and are required to establish an intracellular replication niche.* Infection and immunity, 2006. **74**(12): p. 6965-6972.

107. Boucrot, E., et al., *The intracellular fate of Salmonella depends on the recruitment of kinesin.* Science, 2005. **308**(5725): p. 1174-1178.

108. Henry, T., et al., *The Salmonella effector protein PipB2 is a linker for kinesin-1.* Proceedings of the National Academy of Sciences, 2006. **103**(36): p. 13497-13502.

109. Ohlson, M.B., et al., *Structure and function of Salmonella SifA indicate that its interactions with SKIP, SseJ, and RhoA family GTPases induce endosomal tubulation.* Cell host & microbe, 2008. **4**(5): p. 434-446.

110. Chiu, C.H., T.Y. Lin, and J.T. Ou, *Prevalence of the virulence plasmids of nontyphoid Salmonella in the serovars isolated from humans and their association with bacteremia.* Microbiology and immunology, 1999. **43**(9): p. 899-903.

111. Rotger, R. and J. Casadesús, *The virulence plasmids of Salmonella.* International Microbiology, 2010. **2**(3): p. 177-184.

112. Tezcan-Merdol, D., L. Engstrand, and M. Rhen, *Salmonella enterica SpvB-mediated ADP-ribosylation as an activator for host cell actin degradation.* International journal of medical microbiology, 2005. **295**(4): p. 201-212.

113. Ibarra, J.A. and O. Steele□Mortimer, *Salmonella–the ultimate insider. Salmonella virulence factors that modulate intracellular survival.* Cellular microbiology, 2009. **11**(11): p. 1579-1586.

114. de Jong, H.K., et al., *Host–pathogen interaction in invasive salmonellosis.* PLoS Pathog, 2012. **8**(10).

115. Andino, A. and I. Hanning, *Salmonella enterica: survival, colonization, and virulence differences among serovars.* The Scientific World Journal, 2015. **2015**.

116. Kim, B., et al., *Protecting against antimicrobial effectors in the phagosome allows SodCII to contribute to virulence in Salmonella enterica serovar Typhimurium.* Journal of bacteriology, 2010. **192**(8): p. 2140-2149.

117. Valdez, Y., et al., *Nramp1 expression by dendritic cells modulates inflammatory responses during Salmonella Typhimurium infection.* Cellular microbiology, 2008. **10**(8): p. 1646-1661.

118. Crouch, M.L.V., et al., *Biosynthesis and IroC□dependent export of the siderophore salmochelin are essential for virulence of Salmonella enterica serovar Typhimurium.* Molecular microbiology, 2008. **67**(5): p. 971-983.

119. Santos, R.L., et al., *Life in the inflamed intestine, Salmonella style.* Trends in microbiology, 2009. **17**(11): p. 498-506.

120. Boyer, E., et al., *Acquisition of Mn (II) in addition to Fe (II) is required for full virulence of Salmonella enterica serovar Typhimurium.* Infection and immunity, 2002. **70**(11): p. 6032-6042.

121. Papp-Wallace, K.M., et al., *The CorA Mg2+ channel is required for the virulence of Salmonella enterica serovar typhimurium.* Journal of bacteriology, 2008. **190**(19): p. 6517-6523.

122. Ammendola, S., et al., *High-affinity Zn2+ uptake system ZnuABC is required for bacterial zinc homeostasis in intracellular environments and contributes to the virulence of Salmonella enterica.* Infection and immunity, 2007. **75**(12): p. 5867-5876.

123. Parra-Lopez, C., et al., *A Salmonella protein that is required for resistance to antimicrobial peptides and transport of potassium.* The EMBO journal, 1994. **13**(17): p. 3964.

124. Baggesen, D.L., et al., *Separation of Salmonella typhimurium DT2 and DT135: molecular characterization of isolates of avian origin.* Eur J Epidemiol, 1997. **13**(3): p. 347-52.

125. Wren, B.W., *Microbial genome analysis: insights into virulence, host adaptation and evolution.* Nature Reviews Genetics, 2000. **1**(1): p. 30-39.

126. Kaper, J.B., et al., *Genetics of virulence of enteropathogenic E. coli.* Adv Exp Med Biol, 1997. **412**: p. 279-87.

127. Bäumler, A.J., *The record of horizontal gene transfer in Salmonella.* Trends in microbiology, 1997. **5**(8): p. 318-322.

128. Brüssow, H., C. Canchaya, and W.-D. Hardt, *Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion.* Microbiology and Molecular Biology Reviews, 2004. **68**(3): p. 560-602.

129. Pickard, D., et al., *Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding Salmonella enterica pathogenicity island SPI-7.* Journal of Bacteriology, 2003. **185**(17): p. 5055-5065.

130. Winter, S.E., et al., *Salmonella enterica serovar Typhi conceals the invasion-associated type three secretion system from the innate immune system by gene regulation.* PLoS Pathog, 2014. **10**(7).

131. Song, J., X. Gao, and J.E. Galán, *Structure and function of the Salmonella Typhi chimaeric A2B5 typhoid toxin.* Nature, 2013. **499**(7458): p. 350-354.

132. Gyles, C. and P. Boerlin, *Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease.* Veterinary Pathology Online, 2013. **51**(2): p. 328-340.

133. Parkhill, J., et al., *Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.* Nature, 2001. **413**(6858): p. 848-52.

134. Rychlik, I., et al., *Low-molecular-weight plasmid of Salmonella enterica serovar Enteritidis codes for retron reverse transcriptase and influences phage resistance.* Journal of bacteriology, 2001. **183**(9): p. 2852-2858.

135. Canchaya, C., et al., *Prophage genomics.* Microbiology and Molecular Biology Reviews, 2003. **67**(2): p. 238-276.

136. Cole, S.T., et al., *Massive gene decay in the leprosy bacillus.* Nature, 2001. **409**(6823): p. 1007-11.

137. Fournier, P.-E., et al., *Analysis of the Rickettsia africae genome reveals that virulence acquisition in Rickettsia species may be explained by genome reduction.* BMC genomics, 2009. **10**(1): p. 166.

138. Thomson, N.R., et al., *Chlamydia trachomatis: genome sequence analysis of lymphogranuloma venereum isolates.* Genome research, 2008. **18**(1): p. 161-171.

139. Parkhill, J., et al., *Genome sequence of Yersinia pestis, the causative agent of plague.* Nature, 2001. **413**(6855): p. 523-7.

140. Ventura, M., et al., *Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum.* Microbiology and Molecular Biology Reviews, 2007. **71**(3): p. 495-548.

141. Deng, W., et al., *Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.* Journal of bacteriology, 2003. **185**(7): p. 2330-2337.

142. McClelland, M., et al., *Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid.* Nature genetics, 2004. **36**(12): p. 1268-1274.

143. Efremova, T., et al., *[Invasion of Escherichia coli A2 induces reorganization of actin microfilaments in Hep-2 cells]*. Tsitologiia, 1997. **40**(6): p. 524-528.

144. Finlay, B., S. Ruschkowski, and S. Dedhar, *Cytoskeletal rearrangements accompanying Salmonella entry into epithelial cells.* Journal of Cell Science, 1991. **99**(2): p. 283-296.

145. Hardt, W.-D., et al., *S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells.* Cell, 1998. **93**(5): p. 815-826.

146. Akopyan, K., et al., *Translocation of surface-localized effectors in type III secretion.* Proceedings of the National Academy of Sciences, 2011. **108**(4): p. 1639-1644.

147. Gruenheid, S., et al., *Enteropathogenic E. coli Tir binds Nck to initiate actin pedestal formation in host cells.* Nature cell biology, 2001. **3**(9): p. 856-859.

148. Finlay, B.B. and S. Falkow, *Salmonella interactions with polarized human intestinal Caco-2 epithelial cells.* Journal of Infectious Diseases, 1990. **162**(5): p. 1096-1106.

149. Solano, C., et al., *Virulent strains of Salmonella enteritidis disrupt the epithelial barrier of Caco-2 and HEp-2 cells.* Archives of microbiology, 2001. **175**(1): p. 46-51.

150. Virok, D., et al., *Infection of U937 monocytic cells with Chlamydia pneumoniae induces extensive changes in host cell gene expression.* Journal of Infectious Diseases, 2003. **188**(9): p. 1310-1321.

151. Fitzhenry, R., et al., *Intimin type influences the site of human intestinal mucosal colonisation by enterohaemorrhagic Escherichia coli O157: H7.* Gut, 2002. **50**(2): p. 180-185.

152. Haque, A., et al., *Early interactions of Salmonella enterica serovar typhimurium with human small intestinal epithelial explants.* Gut, 2004. **53**(10): p. 1424-1430.

153. van der Weyden, L., et al., *The mouse genetics toolkit: revealing function and mechanism.* Genome Biol, 2011. **12**(6): p. 224.

154. Platt, R.J., et al., *CRISPR-Cas9 knockin mice for genome editing and cancer modeling.* Cell, 2014. **159**(2): p. 440-55.

155. Seruggia, D. and L. Montoliu, *The new CRISPR–Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals.* Transgenic research, 2014. **23**(5): p. 707-716.

156. Silva, C.A., et al., *Infection of mice by Salmonella enterica serovar Enteritidis involves additional genes that are absent in the genome of serovar Typhimurium.* Infect Immun, 2012. **80**(2): p. 839-49.

157. Vidal, S.M., et al., *Natural resistance to intracellular infections: Nramp1 encodes a membrane phosphoglycoprotein absent in macrophages from susceptible (Nramp1 D169) mouse strains.* The Journal of Immunology, 1996. **157**(8): p. 3559-3568.

158. Fritsche, G., et al., *Slc11a1 (Nramp1) impairs growth of Salmonella enterica serovar typhimurium in macrophages via stimulation of lipocalin-2 expression.* Journal of leukocyte biology, 2012. **92**(2): p. 353-359.

159. Dougan, G., et al., *Immunity to salmonellosis.* Immunological reviews, 2011. **240**(1): p. 196-210.

160. Barthel, M., et al., *Pretreatment of mice with streptomycin provides a Salmonella enterica serovar Typhimurium colitis model that allows analysis of both pathogen and host.* Infect Immun, 2003. **71**(5): p. 2839-58.

161. Hwang, W.Y., et al., *Efficient genome editing in zebrafish using a CRISPR-Cas system.* Nat Biotechnol, 2013. **31**(3): p. 227-9.

162. Takaki, K., et al., *Evaluation of the pathogenesis and treatment of Mycobacterium marinum infection in zebrafish.* Nat Protoc, 2013. **8**(6): p. 1114-24.

163. Swaim, L.E., et al., *Mycobacterium marinum infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity.* Infection and immunity, 2006. **74**(11): p. 6108-6117.

164. Benard, E.L., et al., *Infection of zebrafish embryos with intracellular bacterial pathogens.* J Vis Exp, 2012(61).

165. Meijer, A.H. and H.P. Spaink, *Host-pathogen interactions made transparent with the zebrafish model.* Current drug targets, 2011. **12**(7): p. 1000-1017.

166. Gaiano, N., et al., *Insertional mutagenesis and rapid cloning of essential genes in zebrafish.* Nature 1996. **383**(6603): p. 829-32.

167. Sivasubbu, S., et al., *Insertional mutagenesis strategies in zebrafish.* Genome Biol, 2007. **8**(Suppl 1): p. S9.

168. Wienholds, E. and R.H. Plasterk, *Target-selected gene inactivation in zebrafish.* Methods in cell biology, 2004. **77**: p. 69-90.

169. Kettleborough, R.N., et al., *High-throughput Target-selected Gene Inactivation in Zebrafish.* Methods in cell biology, 2011. **104**: p. 121.

170. Zu, Y., et al., *TALEN-mediated precise genome modification by homologous recombination in zebrafish.* Nature methods, 2013. **10**(4): p. 329-331.

171. Grunwald, D.J., *A revolution coming to a classic model organism.* Nature methods, 2013. **10**(4): p. 303-306.

172. Kettleborough, R.N., et al., *A systematic genome-wide analysis of zebrafish protein-coding gene function.* Nature, 2013. **496**(7446): p. 494-7.

173. Ablain, J., et al., *A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish.* Developmental cell, 2015. **32**(6): p. 756-764.

174. Wu, X., et al., *Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.* Nature biotechnology, 2014. **32**(7): p. 670-676.

175. White, J.K., et al., *Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes.* Cell, 2013. **154**(2): p. 452-64.

176. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proceedings of the National Academy of Sciences, 1977. **74**(11): p. 5088-5090.

177. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proceedings of the National Academy of Sciences, 1990. **87**(12): p. 4576-4579.

178. McQuiston, J., et al., *Molecular phylogeny of the salmonellae: relationships among Salmonella species and subspecies determined from four*

housekeeping genes and evidence of lateral gene transfer events. Journal of bacteriology, 2008. **190**(21): p. 7060-7067.

179. Wu, D., et al., *A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea*. Nature, 2009. **462**(7276): p. 1056-1060.

180. Collins, M., et al., *The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations.* International journal of systematic bacteriology, 1994. **44**(4): p. 812-826.

181. Wong, V.K., et al., *Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter-and intracontinental transmission events.* Nature genetics, 2015. **47**(6): p. 632-639.

182. Ashton, P.M., et al., *Whole Genome Sequencing for the Retrospective Investigation of an Outbreak of Salmonella Typhimurium DT 8.* PLoS currents, 2014. **7**.

183. Mutreja, A., et al., *Evidence for several waves of global transmission in the seventh cholera pandemic.* Nature, 2011. **477**(7365): p. 462-465.

184. Sharp, P.M. and B.H. Hahn, *Origins of HIV and the AIDS pandemic.* Cold Spring Harbor perspectives in medicine, 2011. **1**(1): p. a006841.

185. Maamary, P.G., et al., *Tracing the evolutionary history of the pandemic group A streptococcal M1T1 clone.* The FASEB Journal, 2012. **26**(11): p. 4675-4684.

186. Chewapreecha, C. and S.R. Harris, *Dense genomic sampling identifies highways of pneumococcal recombination.* 2014. **46**(3): p. 305-9.

187. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* Journal of molecular biology, 1975. **94**(3): p. 441-448.

188. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the National Academy of Sciences, 1977. **74**(12): p. 5463-5467.

189. Fleischmann, R.D., et al., *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* Science, 1995. **269**(5223): p. 496-512.

190. Mardis, E.R., *Next-generation sequencing platforms.* Annual review of analytical chemistry, 2013. **6**: p. 287-303.

191. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC Genomics, 2012. **13**: p. 341.

192. Galanis, E., et al., *Web-based surveillance and global Salmonella distribution, 2000–2002.* Emerging infectious diseases, 2006. **12**(3): p. 381.

193. Antony, B., et al., *Food poisoning due to Salmonella enterica serotype Weltevreden in Mangalore.* Indian journal of medical microbiology, 2009. **27**(3): p. 257.

194. Zerbino, D.R., *Using the Velvet de novo assembler for short-read sequencing technologies.* Curr Protoc Bioinformatics, 2010. **Chapter 11**: p. Unit 11 5.

195. Boetzer, M., et al., *Scaffolding pre-assembled contigs using SSPACE.* Bioinformatics, 2011. **27**(4): p. 578-9.

196. Boetzer, M. and W. Pirovano, *Toward almost closed genomes with GapFiller.* Genome Biol, 2012. **13**(6): p. R56.

197. Chin, C.S., et al., *Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.* Nat Methods, 2013. **10**(6): p. 563-9.

198. Sommer, D.D., et al., *Minimus: a fast, lightweight genome assembler.* BMC Bioinformatics, 2007. **8**: p. 64.

199. Hunt, M., et al., *REAPR: a universal tool for genome assembly evaluation.* Genome Biol, 2013. **14**(5): p. R47.

200. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.

201. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification.* BMC Bioinformatics, 2010. **11**: p. 119.

202. Laslett, D. and B. Canback, *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences.* Nucleic Acids Res, 2004. **32**(1): p. 11-6.

203. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p. 421.

204. Wood, D.E. and S.L. Salzberg, *Kraken: ultrafast metagenomic sequence classification using exact alignments.* Genome Biol, 2014. **15**(3): p. R46.

205. Croucher, N.J., et al., *Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.* Nucleic Acids Res, 2015. **43**(3): p. e15.

206. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.* Bioinformatics, 2014. **30**(9): p. 1312-3.

207. Ashkenazy, H., et al., *FastML: a web server for probabilistic reconstruction of ancestral sequences.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W580-4.

208. Kaput, J., et al., *Planning the human variome project: The Spain report.* Human mutation, 2009. **30**(4): p. 496-510.

209. Cheng, L., et al., *Hierarchical and spatially explicit clustering of DNA sequences with BAPS software.* Mol Biol Evol, 2013. **30**(5): p. 1224-8.

210. McLaren, W., et al., *Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor*. Bioinformatics, 2010. **26**(16): p. 2069-70.

211. Hunt, M., Harris, S.R & Mather, A.E, *ARIBA: Antibiotic Resistance Identification By Assembly.* 2015: p. https://github.com/sanger-pathogens/ariba.

212. Fu, L., et al., *CD-HIT: accelerated for clustering the next-generation sequencing data.* Bioinformatics, 2012. **28**(23): p. 3150-2.

213. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. **19**(5): p. 455-77.

214. Inouye, M., et al., *SRST2: Rapid genomic surveillance for public health and hospital microbiology labs.* Genome Med, 2014. **6**(11): p. 90.

215. Kurtz, S., et al., *Versatile and open software for comparing large genomes.* Genome Biol, 2004. **5**(2): p. R12.

216. Page, A.J., et al., *Roary: rapid large-scale prokaryote pan genome analysis.* Bioinformatics, 2015.

217. Enright, A.J., S. Van Dongen, and C.A. Ouzounis, *An efficient algorithm for large-scale detection of protein families.* Nucleic Acids Res, 2002. **30**(7): p. 1575-84.

218. Fouts, D.E., et al., *PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species.* Nucleic Acids Res, 2012. **40**(22): p. e172.

219. McKelvie, N.D., et al., *Expression of heterologous antigens in Salmonella Typhimurium vaccine vectors using the in vivo-inducible, SPI-2 promoter, ssaG.* Vaccine, 2004. **22**(25-26): p. 3243-55.

220. Yu, J., et al., *Interaction of enteric bacterial pathogens with murine embryonic stem cells.* Infect Immun, 2009. **77**(2): p. 585-97.

221. Popoff, M.Y., J. Bockemuhl, and F.W. Hickman-Brenner, *Supplement 1996 (no. 40) to the Kauffmann-White scheme.* Res Microbiol, 1997. **148**(9): p. 811-4.

222. Douce, G.R., Amin, II, and J. Stephen, *Invasion of HEp-2 cells by strains of Salmonella typhimurium of different virulence in relation to gastroenteritis.* J Med Microbiol, 1991. **35**(6): p. 349-57.

223. Malick, L.E. and R.B. Wilson, *Modified thiocarbohydrazide procedure for scanning electron microscopy: routine use for normal, pathological, or experimental tissues.* Stain Technol, 1975. **50**(4): p. 265-9.

224. Richter-Dahlfors, A., A.M. Buchan, and B.B. Finlay, *Murine salmonellosis studied by confocal microscopy: Salmonella typhimurium resides intracellularly inside macrophages and exerts a cytotoxic effect on phagocytes in vivo.* J Exp Med, 1997. **186**(4): p. 569-80.

225. Shiau, C.E., et al., *Differential requirement for irf8 in formation of embryonic and adult macrophages in zebrafish.* PLoS One, 2015. **10**(1): p. e0117513.

226. McClelland, M., et al., *Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.* Nature, 2001. **413**(6858): p. 852-6.

227. Thomson, N.R., et al., *Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways.* Genome Res, 2008. **18**(10): p. 1624-37.

228. Bhowmick, P.P., et al., *Serotyping & molecular characterization for study of genetic diversity among seafood associated nontyphoidal Salmonella serovars.* Indian J Med Res, 2012. **135**: p. 371-81.

229. Ponce, E., et al., *Prevalence and characterization of Salmonella enterica serovar Weltevreden from imported seafood.* Food Microbiol, 2008. **25**(1): p. 29-35.

230. Jain, P., et al., *Salmonella enterica serovar Weltevreden ST1500 associated foodborne outbreak in Pune, India.* Indian J Med Res, 2015. **141**(2): p. 239-41.

231. D'Ortenzio, E., et al., *First report of a Salmonella enterica serovar Weltevreden outbreak on Reunion Island, France, August 2007.* Euro Surveill, 2008. **13**(32).

232. Noor Uddin, G.M., et al., *Clonal Occurrence of Salmonella Weltevreden in Cultured Shrimp in the Mekong Delta, Vietnam.* PLoS One, 2015. **10**(7): p. e0134252.

233. Dunn, J., et al., *Laboratory-based Salmonella surveillance in Fiji, 2004-2005.* Pac Health Dialog, 2005. **12**(2): p. 53-9.

234. Emberland, K.E., et al., *Outbreak of Salmonella Weltevreden infections in Norway, Denmark and Finland associated with alfalfa sprouts, July-October 2007.* Euro Surveill, 2007. **12**(11): p. E071129.4.

235. Patil, A.B., B.V. Krishna, and M.R. Chandrasekhar, *Neonatal sepsis caused by Salmonella enterica serovar Weltevreden.* Southeast Asian J Trop Med Public Health, 2006. **37**(6): p. 1175-8.

236. Desikan, P., et al., *Isolated ulcerative skin lesion caused by Salmonella Weltevreden.* J Infect Dev Ctries, 2009. **3**(7): p. 569-71.

237. Obana, M., et al., *[A fatal case of acute enteritis caused by Salmonella Weltevreden after travel to Indonesia]*. Kansenshogaku Zasshi, 1996. **70**(3): p. 251-4.

238. Brankatschk, K., et al., *Comparative genomic analysis of Salmonella enterica subsp. enterica serovar Weltevreden foodborne strains with other serovars.* Int J Food Microbiol, 2012. **155**(3): p. 247-56.

239. Deekshit, V.K., et al., *Draft Genome Sequence of Multidrug Resistant Salmonella enterica serovar Weltevreden Isolated from Seafood.* J Genomics, 2015. **3**: p. 57-8.

240. Leekitcharoenphon, P., et al., *Genomic variation in Salmonella enterica core genes for epidemiological typing.* BMC Genomics, 2012. **13**: p. 88.

241. Marzel, A., et al., *Integrative analysis of Salmonellosis in Israel reveals association of Salmonella enterica Serovar 9,12:l,v:- with extraintestinal infections, dissemination of endemic S. enterica Serovar Typhimurium DT104 biotypes, and severe underreporting of outbreaks.* J Clin Microbiol, 2014. **52**(6): p. 2078-88.

242. Han, J., et al., *DNA sequence analysis of plasmids from multidrug resistant Salmonella enterica serotype Heidelberg isolates.* PLoS One, 2012. **7**(12): p. e51160.

243. Gilmour, M.W., et al., *The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics.* Plasmid, 2004. **52**(3): p. 182-202.

244. Sumrall, E.T., et al., *Dissemination of the transmissible quinolone-resistance gene qnrS1 by IncX plasmids in Nigeria.* PLoS One, 2014. **9**(10): p. e110279.

245. Wang, J., et al., *Nucleotide sequences of 16 transmissible plasmids identified in nine multidrug-resistant Escherichia coli isolates expressing an ESBL phenotype isolated from food-producing animals and healthy humans.* J Antimicrob Chemother, 2014. **69**(10): p. 2658-68.

246. Contreras-Moreira, B. and P. Vinuesa, *GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis.* Applied and environmental microbiology, 2013. **79**(24): p. 7696-7701.

247. Bullas, L.R., et al., *Salmonella phage PSP3, another member of the P2-like phage group.* Virology, 1992. **188**(1): p. 414.

248. Leung, K.Y. and B.B. Finlay, *Intracellular replication is essential for the virulence of Salmonella typhimurium.* Proc Natl Acad Sci U S A, 1991. **88**(24): p. 11470-4.

249. Chaudhuri, R.R., et al., *Comprehensive identification of Salmonella enterica serovar Typhimurium genes required for infection of BALB/c mice.* PLoS Pathog, 2009. **5**(7): p. e1000529.

250. Zaharik, M.L., et al., *The Salmonella enterica serovar typhimurium divalent cation transport systems MntH and SitABCD are essential for virulence in an Nramp1G169 murine typhoid model.* Infect Immun, 2004. **72**(9): p. 5522-5.

251. Vaas, L.A., et al., *Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics.* PLoS One, 2012. **7**(4): p. e34846.

252. Bar-Peled, M. and M.A. O'Neill, *Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling.* Annu Rev Plant Biol, 2011. **62**: p. 127-55.

253. Jeffries, T.W., *Utilization of xylose by bacteria, yeasts, and fungi.* Adv Biochem Eng Biotechnol, 1983. **27**: p. 1-32.

254. Cossart, P. and P.J. Sansonetti, *Bacterial invasion: the paradigms of enteroinvasive pathogens.* Science, 2004. **304**(5668): p. 242-8.

255. Heinitz, M.L., et al., *Incidence of Salmonella in fish and seafood.* Journal of Food Protection®, 2000. **63**(5): p. 579-592.

256. Sullivan, C. and C.H. Kim, *Zebrafish as a model for infectious disease and immune function.* Fish Shellfish Immunol, 2008. **25**(4): p. 341-50.

257. van der Sar, A.M., et al., *Zebrafish embryos as a model host for the real time analysis of Salmonella typhimurium infections.* Cell Microbiol, 2003. **5**(9): p. 601-11.

258. Menudier, A., F. Rougier, and C. Bosgiraud, *Comparative virulence between different strains of Listeria in zebrafish (Brachydanio rerio) and mice.* Pathologie-biologie, 1996. **44**(9): p. 783-789.

259. Neely, M.N., J.D. Pfeifer, and M. Caparon, *Streptococcus-zebrafish model of bacterial pathogenesis.* Infection and Immunity, 2002. **70**(7): p. 3904-3914.

260. Lin, B., et al., *Acute phase response in zebrafish upon Aeromonas salmonicida and Staphylococcus aureus infection: striking similarities and obvious differences with mammals.* Molecular immunology, 2007. **44**(4): p. 295-301.

261. Prouty, M.G., et al., *Zebrafish-Mycobacterium marinum model for mycobacterial pathogenesis.* FEMS microbiology letters, 2003. **225**(2): p. 177-182.

262. Whipps, C.M., S.T. Dougan, and M.L. Kent, *Mycobacterium haemophilum infections of zebrafish (Danio rerio) in research facilities.* FEMS microbiology letters, 2007. **270**(1): p. 21-26.

263. Rojo, I., et al., *Innate immune gene expression in individual zebrafish after Listonella anguillarum inoculation.* Fish Shellfish Immunol, 2007. **23**(6): p. 1285-93.

264. Pressley, M.E., et al., *Pathogenesis and inflammatory response to Edwardsiella tarda infection in the zebrafish.* Developmental & Comparative Immunology, 2005. **29**(6): p. 501-513.

265. Moyer, T.R. and D.W. Hunnicutt, *Susceptibility of zebra fish Danio rerio to infection by Flavobacterium columnare and F. johnsoniae.* Diseases of aquatic organisms, 2007. **76**(1): p. 39.

266. Davis, J.M., et al., *Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos.* Immunity, 2002. **17**(6): p. 693-702.

267. Parsons, B.N., et al., *Invasive non-typhoidal Salmonella typhimurium ST313 are not host-restricted and have an invasive phenotype in experimentally infected chickens.* PLoS Negl Trop Dis, 2013. **7**(10): p. e2487.

268. Torraca, V., et al., *Macrophage-pathogen interactions in infectious diseases: new therapeutic insights from the zebrafish host model.* Dis Model Mech, 2014. **7**(7): p. 785-97.

269. Kingsley, R.A., et al., *Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype.* Genome Res, 2009. **19**(12): p. 2279-87.

270. Baker, S., et al., *Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission.* Open biology, 2011. **1**(2): p. 110008.

271. Holt, K.E., et al., *High-throughput bacterial SNP typing identifies distinct clusters of Salmonella Typhi causing typhoid in Nepalese children.* BMC infectious diseases, 2010. **10**(1): p. 144.

272. Donadio, S., P. Monciardini, and M. Sosio, *Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics.* Nat Prod Rep, 2007. **24**(5): p. 1073-109.

273. Salvatore, M., et al., *alpha-Defensin inhibits influenza virus replication by cell-mediated mechanism(s).* J Infect Dis, 2007. **196**(6): p. 835-43.

274. Huang, C.M., H.C. Chen, and C.H. Zierdt, *Magainin analogs effective against pathogenic protozoa.* Antimicrob Agents Chemother, 1990. **34**(9): p. 1824-6.

275. Schluesener, H.J., et al., *Leukocytic antimicrobial peptides kill autoimmune T cells.* J Neuroimmunol, 1993. **47**(2): p. 199-202.

276. Hoskin, D.W. and A. Ramamoorthy, *Studies on anticancer activities of antimicrobial peptides.* Biochim Biophys Acta, 2008. **1778**(2): p. 357-75.

277. Lupetti, A., et al., *Antimicrobial peptides: therapeutic potential for the treatment of Candida infections.* Expert Opin Investig Drugs, 2002. **11**(2): p. 309-18.

278. Pushpanathan, M., P. Gunasekaran, and J. Rajendhran, *Antimicrobial peptides: versatile biological properties.* International journal of peptides, 2013. **2013**.

279. Daw, M.A. and F.R. Falkiner, *Bacteriocins: nature, function and structure.* Micron, 1996. **27**(6): p. 467-79.

280. Dye, B.R., et al., *In vitro generation of human pluripotent stem cell derived lung organoids.* Elife, 2015. **4**: p. e05098.

281. Xia, Y., et al., *The generation of kidney organoids by differentiation of human pluripotent cells to ureteric bud progenitor–like cells.* Nature protocols, 2014. **9**(11): p. 2693-2704.

282. Brustle, O., *Developmental neuroscience: Miniature human brains.* Nature, 2013. **501**(7467): p. 319-20.

283. Spence, J.R., et al., *Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro.* Nature, 2011. **470**(7332): p. 105-9.

284. Senju, S., et al., *Generation of dendritic cells and macrophages from human induced pluripotent stem cells aiming at cell therapy.* Gene therapy, 2011. **18**(9): p. 874-883.

285. Hale, C., et al., *Induced Pluripotent Stem Cell Derived Macrophages as a Cellular System to Study Salmonella and Other Pathogens.* PloS one, 2015. **10**(5).

286. Yeung, A., et al., *Conditional-ready mouse embryonic stem cell derived macrophages enable the study of essential genes in macrophage function.* Scientific reports, 2015. **5**: p. 8908.

287. Forbester, J.L., et al., *The interaction of Salmonella enterica Serovar Typhimurium with intestinal organoids derived from human induced pluripotent stem cells.* Infection and immunity, 2015: p. IAI. 00161-15.

288. Toapanta, F.R., et al., *Oral Wild-Type Salmonella Typhi Challenge Induces Activation of Circulating Monocytes and Dendritic Cells in Individuals Who Develop Typhoid Disease.* PLoS Negl Trop Dis, 2015. **9**(6): p. e0003837.

289. Hall, C.J., et al., *Infection-responsive expansion of the hematopoietic stem and progenitor cell compartment in zebrafish is dependent upon inducible nitric oxide.* Cell Stem Cell, 2012. **10**(2): p. 198-209.

290. Liongue, C., et al., *Zebrafish granulocyte colony-stimulating factor receptor signaling promotes myelopoiesis and myeloid cell migration.* Blood, 2009. **113**(11): p. 2535-46.

291. Sood, R., et al., *Development of multilineage adult hematopoiesis in the zebrafish with a runx1 truncation mutation.* Blood, 2010. **115**(14): p. 2806-9.

Appendix 1: Samples' information (metadata)

The table below summarises key information of all samples used in the phylogenetic analysis. This includes the origin (country and region) of the isolates, year and source of isolation and antibiotic resistance profile.

Lane	Name	Country	Region	Year	Source	AMR	Predicted AMR	Cluster	Sub-Cluster
12227_3#4	2011_00503	Guadeloupe	Latin America	2011	Human	Susceptible	Susceptible	1	3
12227_3#10	2011_11351	Guadeloupe	Latin America	2011	Human	Susceptible	Susceptible	1	3
12216_4#54	09_5462	Guadeloupe	Latin America	2009	Human	Susceptible	Susceptible	1	2
12227_3#57	2013_1467	France	Europe	2013	Food (meat and vegetables)	Susceptible	Susceptible	2	4
12227_3#56	2013_1418	France	Europe	2013	Seafood and fish	Susceptible	Susceptible	1	1
12227_3#55	2013_1051	France	Europe	2013	Food (meat and vegetables)	Susceptible	Susceptible	2	4
12227_3#54	2013_1032	France	Europe	2013	Food (meat and vegetables)	Susceptible	Susceptible	1	3
12227_3#53	2013_1005	France	Europe	2013	Food (meat and vegetables)	Susceptible	Susceptible	1	3
12227_3#51	2013_101	France	Europe	2013	Seafood and fish	Susceptible	Susceptible	1	3
12227_3#50	2013_71	France	Europe	2013	Food (meat and vegetables)	Susceptible	Susceptible	1	3
12227_3#44	2013_2776	France	Europe	2013	Food (meat and vegetables)	MDR	Resistant	1	2
12227_3#35	2013_4066	France	Europe	2013	Seafood and fish	Susceptible	Susceptible	2	4
12227_3#28	2006_3866	France	Europe	2006	Seafood and fish	Susceptible	Susceptible	1	3
12227_3#27	2006_3740	France	Europe	2006	Seafood and fish	Susceptible	Susceptible	1	3
12227_3#6	2011_02279	Mayotte Island	Indian Ocean	2011	Human	Ampicillin	Resistant	2	5

12227_3#46	2013_2847	La Reunion Island	Indian Ocean	2013	Animal	Susceptible	Susceptible	2	5
12227_3#41	2012_4716	La Reunion Island	Indian Ocean	2012	Environment	Susceptible	Susceptible	2	5
12227_3#40	2012_4538	La Reunion Island	Indian Ocean	2012	Food (meat and vegetables)	Susceptible	Susceptible	2	5
12227_3#37	2012_2981	La Reunion Island	Indian Ocean	2012	Environment	Susceptible	Susceptible	2	5
12227_3#36	2012_2474	La Reunion Island	Indian Ocean	2012	Animal	Susceptible	Susceptible	2	5
12227_3#34	2013_3667	La Reunion Island	Indian Ocean	2013	Animal	Susceptible	Susceptible	2	5
12227_3#32	2013_3456	La Reunion Island	Indian Ocean	2013	Food (meat and vegetables)	Susceptible	Susceptible	2	5
12227_3#30	2013_3736	La Reunion Island	Indian Ocean	2013	Animal	Susceptible	Susceptible	2	5
12227_3#29	2013_3452	La Reunion Island	Indian Ocean	2013	Food (meat and vegetables)	Susceptible	Susceptible	2	5
12227_3#25	2013_05036	La Reunion Island	Indian Ocean	2013	Human	Susceptible	Susceptible	2	5
12227_3#2	2011_00200	La Reunion Island	Indian Ocean	2011	Human	Susceptible	Susceptible	2	5
12227_3#18	2013_00011	La Reunion Island	Indian Ocean	2013	Human	Susceptible	Susceptible	2	5
12227_3#17	2012_01076	La Reunion Island	Indian Ocean	2012	Human	Susceptible	Susceptible	2	5
12227_3#16	2012_02882	La Reunion Island	Indian Ocean	2012	Human	Susceptible	Susceptible	2	5
12227_3#14	2012_02227	La Reunion Island	Indian Ocean	2012	Human	Susceptible	Susceptible	2	5
12227_3#1	2010_06598	La Reunion Island	Indian Ocean	2010	Human	Susceptible	Susceptible	2	5

12216_4#61	2010_09035	Mayotte Island	Indian Ocean	2010	Human	Susceptible	Susceptible	2	5
12216_4#55	09_8500	Mauritius	Indian Ocean	2009	Human	Susceptible	Susceptible	2	5
12216_4#51	08_2437	Maldives	Indian Ocean	2008	Human	Susceptible	Susceptible	2	4
12216_4#45	00_9879	Madagascar	Indian Ocean	2000	Human	Susceptible	Susceptible	2	5
12216_4#43	99_3134	La Reunion Island	Indian Ocean	1999	Human	Susceptible	Susceptible	2	5
12227_3#21	2013_02479	Algeria	North Africa	2013	Human	Susceptible	Susceptible	1	1
12227_3#7	2011_03604	Tahiti Island	Oceania	2011	Human	Susceptible	Susceptible	2	4
12227_3#49	2013_69	New Caledonia	Oceania	2013	Animal	Susceptible	Susceptible	2	4
12227_3#48	2013_2912	New Caledonia	Oceania	2013	Industrial (porcine feed)	MDR	Susceptible	2	4
12227_3#47	2013_2908	New Caledonia	Oceania	2013	Industrial (porcine feed)	Susceptible	Susceptible	2	4
12227_3#45	2013_2778	New Caledonia	Oceania	2013	Seafood and fish	Susceptible	Susceptible	2	4
12227_3#33	2013_3102	New Caledonia	Oceania	2013	Food (meat and vegetables)	Susceptible	Susceptible	2	4
12227_3#31	2013_3101	New Caledonia	Oceania	2013	Food (meat and vegetables)	Susceptible	Susceptible	2	4
12227_3#3	2011_00324	Tahiti Island	Oceania	2011	Human	Susceptible	Susceptible	2	4
12227_3#23	2013_104810	New Caledonia	Oceania	2013	Human	Susceptible	Susceptible	2	4
12227_3#20	2013_02134	Tahiti Island	Oceania	2013	Human	Susceptible	Susceptible	2	4

12227_3#19	2013_00482	Tahiti Island	Oceania	2013	Human	Susceptible	Susceptible	2	4
12227_3#13	2012_08537	New Caledonia	Oceania	2012	Human	Susceptible	Susceptible	2	4
12227_3#12	2012_01335	Tahiti Island	Oceania	2012	Human	Susceptible	Susceptible	2	4
12227_3#11	2012_05005	Tahiti Island	Oceania	2012	Human	Susceptible	Susceptible	2	4
12216_4#62	2010_08341	Tahiti Island	Oceania	2010	Human	Susceptible	Susceptible	2	4
12216_4#47	03_1986	New Caledonia	Oceania	2003	Human	MDR	Resistant	2	4
12216_4#42	98_11262	New Caledonia	Oceania	1998	Human	Susceptible	Susceptible	2	4
12227_3#8	2011_07037	French Guyana	Latin America	2011	Human	Susceptible	Susceptible	1	2
12227_3#24	2013_04851	French Guyana	Latin America	2013	Human	Susceptible	Susceptible	1	3
12216_4#49	03_5461	French Guyana	Latin America	2003	Human	Susceptible	Susceptible	1	2
12216_4#48	03_4395	French Guyana	Latin America	2003	Human	Susceptible	Susceptible	1	2
12227_3#9	2011_09395	India	South and South-East Asia	2011	Human	Susceptible	Susceptible	1	1
12227_3#43	2013_2518	India	South and South-East Asia	2013	Environment	Susceptible	Susceptible	1	2
12227_3#42	2013_2515	India	South and South-East Asia	2013	Food (meat and vegetables)	Susceptible	Susceptible	2	4

12227_3#39	2012_3614	India	South and South-East Asia	2012	Seafood and fish	Susceptible	Susceptible	1	1
12227_3#38	2012_3395	India	South and South-East Asia	2012	Seafood and fish	Susceptible	Susceptible	2	4
12227_3#15	2012_01497	Thailand	South and South-East Asia	2012	Human	Susceptible	Susceptible	1	3
12216_4#60	2010_08132	India	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	2
12216_4#59	2010_07622	Sri Lanka	South and South-East Asia	2010	Human	Susceptible	Susceptible	2	4
12216_4#41	840K	Sri Lanka	South and South-East Asia	1956	Human	Susceptible	Susceptible	2	4
Pacbio_10259_v0.2	10259	Vietnam	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	3
9472_3#9	30291	Vietnam	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	3
9472_3#8	20510	Vietnam	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	3
9472_3#7	20372	Vietnam	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	2

9472_3#6	20069	Vietnam	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	3
9472_3#5	10347	Vietnam	South and South-East Asia	2009	Human	Susceptible	Resistant	1	2
9472_3#4	10290	Vietnam	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	3
9472_3#34	003_D14	Vietnam	South and South-East Asia	2005	Human	Susceptible	Susceptible	1	1
9472_3#33	C2512	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	2
9472_3#32	C2511	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	1
9472_3#31	C2471	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	3
9472_3#30	C2377	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	3
9472_3#28	C2346	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	1
9472_3#27	C2248	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	3

9472_3#25	C2036	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	3
9472_3#24	C2142	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	1
9472_3#23	170_NVTN	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	3
9472_3#22	008_PNTL	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	3
9472_3#21	179_LTKT	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	1
9472_3#20	194_SL	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	1
9472_3#2	10162	Vietnam	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	2
9472_3#19	184_VTHT	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	2	4
9472_3#18	46_DNBH	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	3
9472_3#17	38_NTMD	Vietnam	South and South-East Asia	2007	Human	Susceptible	Resistant	1	2

9472_3#16	8_LTPO	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	3
9472_3#15	132_NTNK	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	3
9472_3#14	94_VNQN	Vietnam	South and South-East Asia	2007	Human	Susceptible	Resistant	1	3
9472_3#13	106_MR	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	1
9472_3#12	63_AYSA	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	1
9472_3#11	72_LNPT	Vietnam	South and South-East Asia	2007	Human	Susceptible	Susceptible	1	3
9472_3#10	30438	Vietnam	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	1
9472_3#1	iNT_635	Vietnam	South and South-East Asia	2009	Human	Susceptible	Resistant	1	2
12227_3#5	2011_00823	Indonesia	South and South-East Asia	2011	Human	Susceptible	Susceptible	1	1
12227_3#26	2013_05421	Thailand	South and South-East Asia	2013	Human	Susceptible	Susceptible	1	3

12227_3#22	2013_03357	Thailand	South and South-East Asia	2013	Human	Susceptible	Susceptible	2	4
12216_4#58	2010_08825	Malaysia	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	2
12216_4#57	2010_09500	Laos	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	3
12216_4#56	2010_05280	Indonesia	South and South-East Asia	2010	Human	Susceptible	Susceptible	1	1
12216_4#53	09_4703	Indonesia	South and South-East Asia	2009	Human	Susceptible	Susceptible	1	1
12216_4#46	02_1171	Thailand	South and South-East Asia	2002	Human	Susceptible	Susceptible	2	4
12216_4#44	00_9824	Thailand	South and South-East Asia	2000	Human	Susceptible	Susceptible	1	3
12216_4#40	139K	Indonesia	South and South-East Asia	1940	Human	Susceptible	Susceptible	1	1
10900_1#32	SR046	Vietnam	South and South-East Asia	NA	Animal	Susceptible	Susceptible	1	3
10900_1#31	SR134	Vietnam	South and South-East Asia	NA	Animal	Susceptible	Susceptible	1	1

10868_1#91	74_V_379	Vietnam	South and South-East Asia	NA	Animal	Susceptible	Susceptible	1	3
10868_1#90	7_H_437	Vietnam	South and South-East Asia	NA	Animal	Susceptible	Susceptible	1	3
10868_1#89	71_V_366	Vietnam	South and South-East Asia	NA	Animal	Susceptible	Susceptible	1	3
12227_3#52	2013_690	Unknown	Unknown	2013	Seafood and fish	Susceptible	Susceptible	2	4

Appendix 2: Illumina HiSeq output

The table below summarises the outcome of the Illumina HiSeq runs of all samples included in the study.

		Mapping		*De novo* assembly			
Name	**Reads**	**Mapped %**	**Coverage**	**Length**	**Contigs**	**N50**	**Genes**
10868_1#89	3889532	98.1	73.94	5044941	97	112822	4949
10868_1#90	4688928	98.7	89.64	5006424	78	144704	4905
10868_1#91	4260526	93.1	76.88	5044922	177	147001	4919
10900_1#31	2801512	95.5	51.83	5151860	80	143020	5053
10900_1#32	3599042	99.4	69.34	5045168	76	149123	4951
12216_4#40	5144766	99	98.65	4868524	71	134434	4748
12216_4#41	4699890	98.4	89.61	4859440	65	144627	4757
12216_4#42	4363564	98.5	83.26	4906994	64	133432	4769
12216_4#43	4341900	94.4	79.43	5007548	57	214657	4946
12216_4#44	4519078	99.5	87.12	5004210	78	149409	4894
12216_4#45	4708460	97.7	89.15	5011241	63	164965	4917
12216_4#46	4592746	98.2	87.41	4944440	74	139950	4871
12216_4#47	4703734	84.5	76.97	5005467	62	172717	4948
12216_4#48	4753954	99.2	91.38	4952303	72	161583	4828
12216_4#49	4917208	99.2	94.49	4842349	69	175797	4722
12216_4#51	4428042	97.9	84	4966191	60	195013	4857
12216_4#53	4566100	98.4	87.07	4895076	72	162400	4784
12216_4#54	5162754	97.9	97.93	5020258	67	195411	4933
12216_4#55	4871450	95.8	90.43	5072792	52	284108	4980

12216_4#56	5284370	97.4	99.69	5052883	88	146745	4960
12216_4#57	4568892	94.5	83.68	5090761	83	149007	4987
12216_4#58	3952184	98.4	75.33	4976493	67	161403	4871
12216_4#59	4459446	98.5	85.1	4921713	50	195414	4790
12216_4#60	4151724	98.9	79.58	4960208	57	195403	4831
12216_4#61	4393012	97.5	83	5007607	55	195318	4916
12216_4#62	4618592	94.3	84.36	4945027	49	251959	4840
12227_3#1	5194158	98.5	99.12	4962601	50	215020	4842
12227_3#10	4570594	98.1	86.89	5042864	79	146957	4934
12227_3#11	4926752	99.1	94.61	4751595	47	238100	4605
12227_3#12	4813016	98.3	91.62	4924337	50	251914	4795
12227_3#13	5236316	98.9	100.31	4915900	55	159187	4771
12227_3#14	5412332	95.8	100.44	5101676	58	214999	4997
12227_3#15	5730108	99.2	110.09	5042706	76	189557	4943
12227_3#16	6052146	97.8	114.67	4996706	56	261708	4900
12227_3#17	5260846	97.4	99.3	5008621	58	143289	4908
12227_3#18	5372556	97.4	101.37	4988534	71	195207	4869
12227_3#19	5187148	98	98.48	4883424	50	212167	4755
12227_3#2	5011300	98.3	95.47	4965443	54	214985	4852
12227_3#20	4480874	98.5	85.54	4923384	48	195388	4780
12227_3#21	4586082	98.1	87.16	4978574	83	146469	4857
12227_3#22	5036370	95.5	93.18	5035718	69	195212	4909
12227_3#23	5827672	98.6	111.31	4944434	57	159186	4818
12227_3#24	5502848	99.6	106.16	5023404	81	149196	4907
12227_3#25	5068292	95.7	94.01	5101214	59	296488	5009
12227_3#26	4538930	99.1	87.17	5038836	67	167656	4945
12227_3#27	5409458	99.5	104.31	5047209	73	149174	4959

12227_3#28	5104036	99.5	98.35	5048510	76	146959	4953
12227_3#29	5053162	97.7	95.65	4976937	55	195293	4860
12227_3#3	5867244	98	111.44	4851166	52	251953	4690
12227_3#30	4883262	98.5	93.17	4965221	48	214747	4847
12227_3#31	5537958	98.3	105.48	4946827	58	159194	4816
12227_3#32	5310382	97.8	100.65	4865035	47	195395	4764
12227_3#33	4583850	98.3	87.29	4947501	64	159222	4819
12227_3#34	4717704	96	87.75	5068657	56	255910	4958
12227_3#35	5417350	99.1	104	4891302	56	162271	4755
12227_3#36	5565292	97.7	105.32	5019360	58	214869	4915
12227_3#37	5349784	97.8	101.4	4995518	59	255919	4890
12227_3#38	4578170	98.4	87.26	4917439	44	282900	4786
12227_3#39	5412292	97.9	102.62	4955321	69	162412	4847
12227_3#4	5644922	98	107.2	5045161	69	175809	4929
12227_3#40	5413274	97.6	102.32	5010535	58	195289	4924
12227_3#41	4807198	90.9	84.64	5226277	59	195402	5169
12227_3#42	5387826	97.8	102.04	4905900	47	236046	4774
12227_3#43	5302882	98.4	101.13	4920067	60	175603	4792
12227_3#44	5684572	97.2	107.07	5089435	79	176985	4997
12227_3#45	5291962	98.2	100.69	4946756	58	227814	4822
12227_3#46	5332544	97.6	100.78	5007769	56	205697	4919
12227_3#47	5327502	98.7	101.85	4926290	52	206554	4775
12227_3#48	4881750	98.6	93.29	4926529	53	236317	4780
12227_3#49	4881894	98.1	92.82	4986584	56	160361	4868
12227_3#5	6169240	98.7	117.95	4954924	75	162452	4808
12227_3#50	4635114	99.5	89.37	5007888	72	149266	4887
12227_3#51	4746446	99.5	91.5	5048359	76	149111	4955

12227_3#52	5429082	98.2	103.28	4965012	51	201132	4850
12227_3#53	4376474	97.3	82.47	5133374	83	192470	5036
12227_3#54	4530426	97.2	85.3	5130868	87	146957	5022
12227_3#55	5260556	97.3	99.17	4927651	50	256825	4811
12227_3#56	5363480	96.6	100.36	5132332	79	181687	5031
12227_3#57	4621298	96.2	86.09	4977137	55	251917	4858
12227_3#6	5679188	94.4	103.83	5098879	66	195247	5018
12227_3#7	6293382	98.3	119.83	4841149	57	176642	4704
12227_3#8	6262826	97.2	117.96	5001905	60	195413	4902
12227_3#9	5228332	98.4	99.63	4991840	87	146861	4875
9472_3#1	12568808	96.3	234.35	4935559	66	195369	4858
9472_3#10	9411182	96.6	176.06	5135834	80	146736	5055
9472_3#11	10538574	99.5	203.19	5052882	76	175982	4976
9472_3#12	9721890	96.9	182.41	5165454	81	146725	5088
9472_3#13	10502146	97.4	198.08	5161880	86	198510	5095
9472_3#14	9740586	99	186.89	5066549	74	175906	4997
9472_3#15	9668256	99.7	186.76	5012030	78	147037	4921
9472_3#16	9882038	98.8	189.1	5058896	74	150539	4991
9472_3#17	9964280	96.3	185.89	5029157	74	175836	4934
9472_3#18	8627588	99.5	166.34	5045229	92	175630	4957
9472_3#19	9728464	97.6	183.97	4934937	54	215006	4818
9472_3#2	9987104	97.9	189.32	4988511	70	162584	4888
9472_3#20	10661266	98.7	203.89	5015618	85	143410	4919
9472_3#21	10592012	96.9	198.87	5169321	86	145882	5086
9472_3#22	9617680	99.3	185.04	5014284	75	162533	4920
9472_3#23	9213008	99.4	177.34	5064218	87	146103	4996
9472_3#24	10013626	97	188.12	5166931	92	134459	5079

9472_3#25	11974942	99.4	230.49	5020379	70	147031	4938
9472_3#27	9596238	99.8	185.5	5005406	72	146959	4899
9472_3#28	8842562	96.9	166.04	5148340	87	146762	5056
9472_3#30	8930070	89.8	154.96	5010158	73	175874	4916
9472_3#31	8948484	98.3	170.44	5085083	65	195350	4995
9472_3#32	10228588	94.4	187.04	5197912	101	134125	5125
9472_3#33	11191750	90.5	195.31	5039988	69	243732	4975
9472_3#34	7232246	87.2	121.62	5160889	79	146730	5074
9472_3#4	8464880	98.6	161.76	5003613	73	146959	4963
9472_3#5	10708366	97.8	202.92	5042617	75	195368	4974
9472_3#6	8478962	98.2	161.31	5087212	67	175850	5003
9472_3#7	9953400	97.4	187.83	4939069	70	162515	4850
9472_3#8	9396910	99.3	180.81	5044982	77	147011	4959
9472_3#9	11845904	99.5	228.27	4998331	75	162227	4890

Appendix 3: SNPs defining the major phylogenetic clusters

A hundred and twelve SNPs were found to enable to discriminate between the 2 major clusters. The table below provides a comprehensive description of each SNP, their position in the genome and putative functions of the genes they were found in.

Coordinates	Continental	Islands	Type	Change	Name	Function
38110	T	G	Nonsynonymous	M1I	yhcR	secreted 5'-nucleotidase
53518	C	G	Nonsynonymous	V1M	ribF	riboflavin biosynthesis protein RibF
150218	T	C	Synonymous	287R	murD	UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase
159866	G	A	Intergenic			
165171	A	G	Synonymous	48V	lysR_1	LysR family transcriptional regulator
171185	G	C	Nonsynonymous	Y110H	hofB	protein transport protein HofB
186542	A	G	Intergenic			
455296	A	G	Intergenic			
482321	T	C	Nonsynonymous	T116A	yajI	lipoprotein
493751	G	A	Intergenic			
551959	T	A	Nonsynonymous	D132N	ybaN	Inner membrane protein YbaN
553182	G	T	Intergenic			
555948	G	A	Nonsynonymous	R136C	SBOV4431	chaperone protein HtpG
623261	A	G	Intergenic			
666014	A	G	Intergenic			
667137	A	G	Intergenic			

694039	C	A	Nonsynonymous	A14V	entF	enterobactin synthetase component F
731339	G	A	Nonsynonymous	N142K	citF_2	citrate lyase subunit alpha
877723	A	G	Nonsynonymous	L16P	sdcS_1	cation transporter
882274	C	T	Nonsynonymous	N160T	10259_00846	membrane protein
885476	T	C	Synonymous	245N	gpmA	phosphoglyceromutase
928510	C	T	Synonymous	217L	ybhL_1	membrane protein
969008	A	G	Intergenic			
1087552	T	C	Synonymous	94D	cydD	cysteine/glutathione ABC transporter membrane/ATP-binding component
1102354	C	T	Nonsynonymous	L171F	dmsB_2	anaerobic dimethyl sulfoxide reductase subunit B
1151458	C	T	Synonymous	268T	10259_01111	amino acid:proton symporter
1202670	T	A	Nonsynonymous	S177P	pepN	aminopeptidase N
1283933	T	C	Synonymous	70P	10259_01257	hypothetical protein
1461980	A	T	Intergenic			
1539870	A	G	Intergenic			
1560429	T	C	Intergenic			

1625658	G	A	Intergenic			
1669118	G	A	Intergenic			
1694992	A	C	Intergenic			
1697445	G	A	Nonsynonymous	P186L	SBOV16411	putative inner membrane protein
1703784	G	C	Nonsynonymous	L19R	10259_01695	protein ydcJ
1720617	T	C	Synonymous	76N	gatC_1	phosphotransferase enzyme
1735902	C	A	Nonsynonymous	V190I	ydcR_1	GntR family transcriptional regulator
1743929	A	G	Synonymous	37R	10259_01734	ssrAB activated gene
1782320	G	A	Intergenic			
1792507	C	T	Synonymous	243Y	galS_1	transcriptional regulator
1827483	T	G	Intergenic			
1842679	C	T	Nonsynonymous	P191L	10259_01827	lipoprotein
1904277	G	A	Synonymous	33P	ydhJ	multidrug resistance efflux pump
1938526	G	T	Nonsynonymous	Y191H	sseC	pathogenicity island 2 effector protein SseC
1945318	G	C	Nonsynonymous	A206T	ssrA	sensor kinase

1948148	C	T	Nonsynonymous	P212A	ycgE_1	MerR family transcriptional regulator
1952986	T	C	Nonsynonymous	N234H	ttrB	tetrathionate reductase subunit B
1955378	C	T	Synonymous	391L	ttrA	tetrathionate reductase subunit A
2072991	C	T	Synonymous	395H	dosC	diguanylate cylase
2104434	A	T	Intergenic			
2155215	T	C	Synonymous	29G	mnmA	tRNA-specific 2-thiouridylase MnmA
2179555	A	C	Nonsynonymous	I24T	ycfS	LD-transpeptidase YcfS
2306947	T	G	Nonsynonymous	I253V	ackA_1	propionate kinase
2312418	C	T	Nonsynonymous	V262L	dacD	penicillin-binding protein
2327448	G	A	Synonymous	20L	hisC	histidinol-phosphate aminotransferase
2349279	C	T	Nonsynonymous	G266D	rfbD_2	dTDP-4-dehydrorhamnose reductase
2368620	T	C	Nonsynonymous	P271H	wcaC	glycosyltransferase
2388569	G	A	Nonsynonymous	G3R	yegN	RND family transporter protein
2397768	A	G	Intergenic			
2487266	C	T	Nonsynonymous	N31H	yeiO	sugar efflux transporter
2571886	C	A	Synonymous	233G	arnB	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase
2644752	C	T	Nonsynonymous	A315V	folC	folylpolyglutamate synthase

2655437	C	A	Intergenic			
2745065	A	G	Synonymous	152R	xapA	purine nucleoside phosphorylase
2751842	T	C	Intergenic			
2806294	A	C	Nonsynonymous	T319I	dapE	succinyl-diaminopimelate desuccinylase
2830342	A	G	Nonsynonymous	I331S	ppx	exopolyphosphatase
2918022	A	C	Nonsynonymous	H332D	dmsA_4	putative anaerobic dimethylsulfoxide reductase
2939599	T	C	Nonsynonymous	L349Q	10259_02939	reductase
3029090	G	A	Nonsynonymous	L36P	nadB	L-aspartate oxidase
3044150	G	A	Synonymous	372K	kgtP	alpha-ketoglutarate transporter
3165103	G	T	Nonsynonymous	G366S	gabR	DeoR family transcriptional regulator
3192557	G	A	Synonymous	87L	srlA	Glucitol/sorbitol permease IIC component
3247227	A	G	Synonymous	216Q	spaR	virulence associated secretory protein
3250702	G	A	Synonymous	12Q	spaI	secretory apparatus ATP synthase (associated with virulence)
3258960	A	C	Intergenic			
3323082	G	A	Nonsynonymous	W394C	SBOV30001	conserved hypothetical protein
3469564	A	G	Synonymous	84Q	yqgD	inner membrane protein

3502405	T	C	Nonsynonymous	R422L	10259_03509	FIC domain-containing protein
3559237	G	T	Nonsynonymous	E431Q	STY3343	putative exported protein
3638294	T	C	Nonsynonymous	R451L	tdcD	propionate/acetate kinase
3710463	G	A	Synonymous	122L	mtgA	monofunctional biosynthetic peptidoglycan transglycosylase
3715222	T	G	Intergenic			
3734518	G	A	Intergenic			
3784257	T	C	Synonymous	260V	acrF	acriflavin resistance protein F
3798695	G	A	Synonymous	51P	fmt	methionyl-tRNA formyltransferase
3844549	C	T	Intergenic			
3861654	A	G	Synonymous	68*	aroB	3-dehydroquinate synthase
3933453	A	G	Intergenic			
3935103	G	A	Synonymous	226S	php	phosphotriesterase
4022010	T	A	Nonsynonymous	A49V	10259_04023	putative inner membrane protein
4054900	G	A	Nonsynonymous	D498G	xylR	xylose operon regulatory protein
4087478	C	A	Intergenic			

4151451	A	G	Intergenic			
4179949	T	C	Synonymous	81N	10259_04171	putative secreted protein
4183764	C	A	Intergenic			
4202714	G	A	Nonsynonymous	A51T	dsdX	permease
4231325	T	G	Nonsynonymous	M5118T	10259_04223	2-oxo-3-deoxygalactonate kinase
4343874	A	G	Nonsynonymous	S55R	hemC	porphobilinogen deaminase
4464873	G	A	Nonsynonymous	L550Q	siaT_2	integral membrane transport protein
4469553	A	C	Nonsynonymous	A561E	cpxA	two-component sensor kinase protein
4478528	G	A	Synonymous	175K	yicJ_2	sodium:galactoside symporter
4525097	G	A	Intergenic			
4528083	T	C	Intergenic			
4537997	A	G	Intergenic			
4613302	A	G	Nonsynonymous	T68A	10259_04597	histidine biosynthesis protein
4651837	A	T	Intergenic			
4659202	C	T	Synonymous	224N	bepC	type-I secretion protein

4676493	T	C	Nonsynonymous	N697H	10259_04641	Ig domain-containing protein
4740492	T	C	Intergenic			
4783210	C	T	Nonsynonymous	P7S	sugE	SugE protein
4785518	A	C	Nonsynonymous	H71Y	frdB	fumarate reductase, iron-sulfur protein
4794271	C	T	Nonsynonymous	S75T	psd	phosphatidylserine decarboxylase proenzyme
4863047	C	T	Intergenic			
4865401	G	A	Nonsynonymous	F78I	iolB	5-deoxyglucuronate isomerase
4867756	G	T	Intergenic			
4867800	G	A	Intergenic			
4872126	T	C	Synonymous	359D	10259_04826	lysosomal glucosyl ceramidase
4883973	G	A	Synonymous	37E	pmbA	peptidase PmbA
4901727	G	A	Nonsynonymous	G84D	mgtA	magnesium-transporting ATPase MgtA
4982215	C	G	Nonsynonymous	R88Q	mdtM	sugar transport protein

Appendix 4: MLST data

The table below summarises the sequence type (ST) and the alleles numbers for the house keeping genes used to compile the MLST profile of all samples used in the study and the ones that were discarded.

Isolate-LANE ID	ST	aroC	dnaN	hemD	hisD	purE	sucA	thrA
10868_1#89.contigs_velvet	1500	130	97	25	125	422	9	101
10868_1#90.contigs_velvet	1500	130	97	25	125	422	9	101
10868_1#91.contigs_velvet	1500	130	97	25	125	422	9	101
10900_1#31.contigs_velvet	1500	130	97	25	125	422	9	101
10900_1#32.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#40.contigs_velvet	559	130	97	25	125	U	9	101
12216_4#41.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#42.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#43.contigs_velvet	559	U	97	25	125	U	9	101
12216_4#44.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#45.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#46.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#47.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#48.contigs_velvet	1500	130	97	25	125	422	9	101

12216_4#50.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#49.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#51.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#52.contigs_velvet	518	101	41	40	184	76	90	3
12216_4#53.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#54.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#55.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#56.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#57.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#58.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#59.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#60.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#61.contigs_velvet	1500	130	97	25	125	422	9	101
12216_4#62.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#10.contigs_velvet	1500	130	97	25	125	422	9	101

12227_3#11.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#12.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#13.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#14.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#15.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#16.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#17.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#18.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#19.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#1.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#20.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#22.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#21.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#24.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#23.contigs_velvet	1500	130	97	25	125	422	9	101

12227_3#26.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#25.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#28.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#27.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#29.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#2.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#30.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#31.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#32.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#33.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#34.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#35.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#36.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#37.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#38.contigs_velvet	559	130	97	25	125	422	9	U

12227_3#39.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#3.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#40.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#41.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#42.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#43.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#44.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#45.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#46.contigs_velvet	559	U	97	25	125	422	9	101
12227_3#47.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#48.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#49.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#4.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#50.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#51.contigs_velvet	1500	130	97	25	125	422	9	101

12227_3#52.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#53.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#54.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#55.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#56.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#57.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#5.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#6.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#7.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#8.contigs_velvet	1500	130	97	25	125	422	9	101
12227_3#9.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#10.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#11.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#12.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#13.contigs_velvet	1500	130	97	25	125	422	9	101

9472_3#14.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#15.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#16.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#17.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#18.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#19.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#1.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#20.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#21.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#22.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#23.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#24.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#25.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#26.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#28.contigs_velvet	1500	130	97	25	125	422	9	101

9472_3#27.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#2.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#29.contigs_velvet	1684	U	U	U	U	281	U	
9472_3#30.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#31.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#32.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#33.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#3.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#34.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#4.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#5.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#6.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#7.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#8.contigs_velvet	1500	130	97	25	125	422	9	101
9472_3#9.contigs_velvet	1500	130	97	25	125	422	9	101

Appendix 5: Phage search in Phast

The table below shows an the outcome of the phage search performed on Phast for 4 *S.* Weltevreden isolates.

***S.* Weltevreden 10259 contig 1 undefined product 1:5062936 forward .5062936, GC%: 52.16%, length = 5062936 bps**

Total : 13 prophage regions have been identified, of which 9 regions are intact, 3 regions are incomplete, 1 regions are questionable.

REGION	REGION_LENGTH	COMPLETENESS	SCORE	#CDS	REGION_POSITION	POSSIBLE PHAGE	GC %	DETAIL
1	41Kb	intact	150	53	624585-665599	PHAGE_Salmon_Fels_1_NC_010391, ...	51.97%	Detail
2	43.8Kb	intact	150	52	1014619-1058492	PHAGE_Salmon_ST64B_NC_004313, ...	49.45%	Detail
3	32.9Kb	questionable	70	9	1057625-1090615	PHAGE_Cronob_vB_CsaM_GAP32_NC_019401, ...	53.59%	Detail
4	47.3Kb	intact	150	57	1153517-1200901	PHAGE_Gifsy_2_NC_010393, ...	51.04%	Detail
5	11.8Kb	incomplete	40	12	1431061-1442946	PHAGE_Entero_HK106_NC_019768, ...	51.78%	Detail
6	47.8Kb	intact	150	63	2107247-2155075	PHAGE_Gifsy_1_NC_010392, ...	51.83%	Detail
7	40.7Kb	intact	105	58	2674176-2714885	PHAGE_Salmon_c341_NC_013059, ...	47.29%	Detail
8	58.5Kb	intact	105	57	2822899-2881489	PHAGE_Escher_TL_2011b_NC_019445, ...	51.06%	Detail
9	18.4Kb	incomplete	50	21	2980916-2999341	PHAGE_Aggreg_S1249_NC_013597, ...	50.88%	Detail
10	37.6Kb	intact	100	30	2987821-3025476	PHAGE_Salmon_SPN3UB_NC_019545, ...	51.98%	Detail
11	34.8Kb	intact	128	45	3060402-3095234	PHAGE_Entero_PsP3_NC_005340, ...	51.90%	Detail
12	24.8Kb	incomplete	30	14	4930004-4954832	PHAGE_Entero_P4_NC_001609, ...	49.57%	Detail
13	7.5Kb	intact	110	15	4943648-4951226	PHAGE_Shigel_SfIV_NC_022749, ...	49.24%	Detail

***S.* Weltevreden C2346 Contig 1 5129845 forward .5129845, GC%: 52.18%, length = 5129845 bps**

Total: 11 prophage regions have been identified, of which 8 regions are intact, 3 regions are incomplete, 0 regions are questionable.

REGION	REGION_LENGTH	COMPLETENESS	SCORE	#CDS	REGION_POSITION	POSSIBLE PHAGE	GC %	DETAIL
1	41Kb	intact	150	54	634260-675274	PHAGE_Salmon_Fels_1_NC_010391, ...	51.97%	Detail
2	26.8Kb	incomplete	60	9	1022187-1049025	PHAGE_Cronob_vB_CsaM_GAP32_NC_019401, ...	53.51%	Detail
3	48.7Kb	intact	150	60	1121849-1170548	PHAGE_Gifsy_2_NC_010393, ...	51.04%	Detail
4	48.1Kb	intact	150	61	1459559-1507668	PHAGE_Gifsy_1_NC_010392, ...	51.80%	Detail
5	19.3Kb	incomplete	40	12	2157246-2176555	PHAGE_Entero_HK106_NC_019768, ...	45.84%	Detail
6	54.4Kb	intact	110	56	2762040-2816527	PHAGE_Entero_phiV10_NC_007804, ...	51.36%	Detail
7	44.5Kb	intact	150	54	2910719-2955279	PHAGE_Salmon_SPN3UB_NC_019545, ...	51.18%	Detail
8	34.8Kb	intact	128	43	2990205-3025037	PHAGE_Entero_PsP3_NC_005340, ...	51.90%	Detail
9	64.6Kb	intact	150	58	4650120-4714751	PHAGE_Salmon_ST64B_NC_004313, ...	51.83%	Detail
10	24.8Kb	incomplete	30	14	4996913-5021741	PHAGE_Entero_	49.57%	Detail

						P4_NC_001609, ...		
11	7.5Kb	intact	110	15	5010557-5018135	PHAGE_Shigel_SfIV_NC_022749, ...	49.24%	Detail

S. Weltevreden 98_11262 Contig 1 4897142, GC%: 52.18%, length = 4897142 bps

Total : 8 prophage regions have been identified, of which 4 regions are intact, 3 regions are incomplete, 1 regions are questionable.

REGION	REGION_LENGTH	COMPLETENESS	SCORE	#CDS	REGION_POSITION	POSSIBLE PHAGE	GC %	DETAIL
1	51.6Kb	intact	120	79	940-52580	PHAGE_Entero_ST64T_NC_004348, ...	47.13%	Detail
2	6.3Kb	questionable	90	13	792695-799011	PHAGE_Ralsto_phiRSA1_NC_009382, ...	48.41%	Detail
3	12.1Kb	incomplete	40	17	801306-813443	PHAGE_Entero_P4_NC_001609, ...	50.81%	Detail
4	34.1Kb	intact	110	41	2133831-2168020	PHAGE_Vibrio_8_NC_022747, ...	50.01%	Detail
5	48.1Kb	intact	150	62	3446723-3494831	PHAGE_Gifsy_1_NC_010392, ...	51.80%	Detail
6	19.3Kb	incomplete	40	12	4142117-4161426	PHAGE_Entero_HK106_NC_019768, ...	45.84%	Detail
7	48.3Kb	intact	150	58	4391859-4440190	PHAGE_Gifsy_2_NC_010393, ...	51.09%	Detail
8	18.9Kb	incomplete	60	9	4513062-4532052	PHAGE_Cronob_	53.58%	Detail

						vB_CsaM_GAP32_NC_019401, ...		

S. Weltevreden 99_3134 Contig 1 4977779, GC%: 52.15%, length = 4977779 bps

Total : 11 prophage regions have been identified, of which 7 regions are intact, 3 regions are incomplete, 1 regions are questionable.

REGION	REGION_LENGTH	COMPLETENESS	SCORE	#CDS	REGION_POSITION	POSSIBLE PHAGE	GC %	DETAIL
1	34.9Kb	intact	150	48	309-35286	PHAGE_Salmon_Fels_1_NC_010391, ...	47.34%	Detail
2	51.5Kb	intact	100	54	370802-422363	PHAGE_Entero_P22_NC_002371, ...	48.13%	Detail
3	6.3Kb	questionable	80	12	1162475-1168791	PHAGE_Ralsto_phiRSA1_NC_009382, ...	48.41%	Detail
4	23.5Kb	incomplete	30	14	1158855-1182435	PHAGE_Entero_P4_NC_001609, ...	49.37%	Detail
5	34.3Kb	intact	100	39	2503660-2537985	PHAGE_Vibrio_8_NC_022747, ...	50.01%	Detail
6	34.7Kb	intact	125	48	3050986-3085752	PHAGE_Entero_PsP3_NC_005340, ...	52.60%	Detail
7	61.9Kb	intact	150	56	3861593-3923505	PHAGE_Gifsy_1_NC_010392, ...	51.81%	Detail
8	24.8Kb	incomplete	50	8	3911226-3936104	PHAGE_Cronob_vB_CsaM_GAP32_NC_019401, ...	51.64%	Detail

9	46.1Kb	intact	150	57	4009349-4055531	PHAGE_Gifsy_2_NC_010393, ...	50.73%	Detail
10	11.8Kb	incomplete	40	12	4279868-4291753	PHAGE_Entero_HK106_NC_019768, ...	51.79%	Detail
11	26.8Kb	intact	150	41	4950836-4977691	PHAGE_Gifsy_1_NC_010392, ...	49.80%	Detail

Legend:
REGION: the number assigned to the region
REGION_LENGTH: the length of the sequence of that region (in bp)
COMPLETENESS: a prediction of whether the region contains a intact or incomplete prophage based on the above criteria
SCORE: the score of the region based on the above criteria
#CDS: the number of coding sequnce
REGION_POSITION: the start and end positions of the region on the bacterial chromosome
PHAGE: the phage with the highest number of proteins most similar to those in the region
GC_PERCENTAGE: the percentage of gc nucleotides of the region
DETAIL: detail info of the region

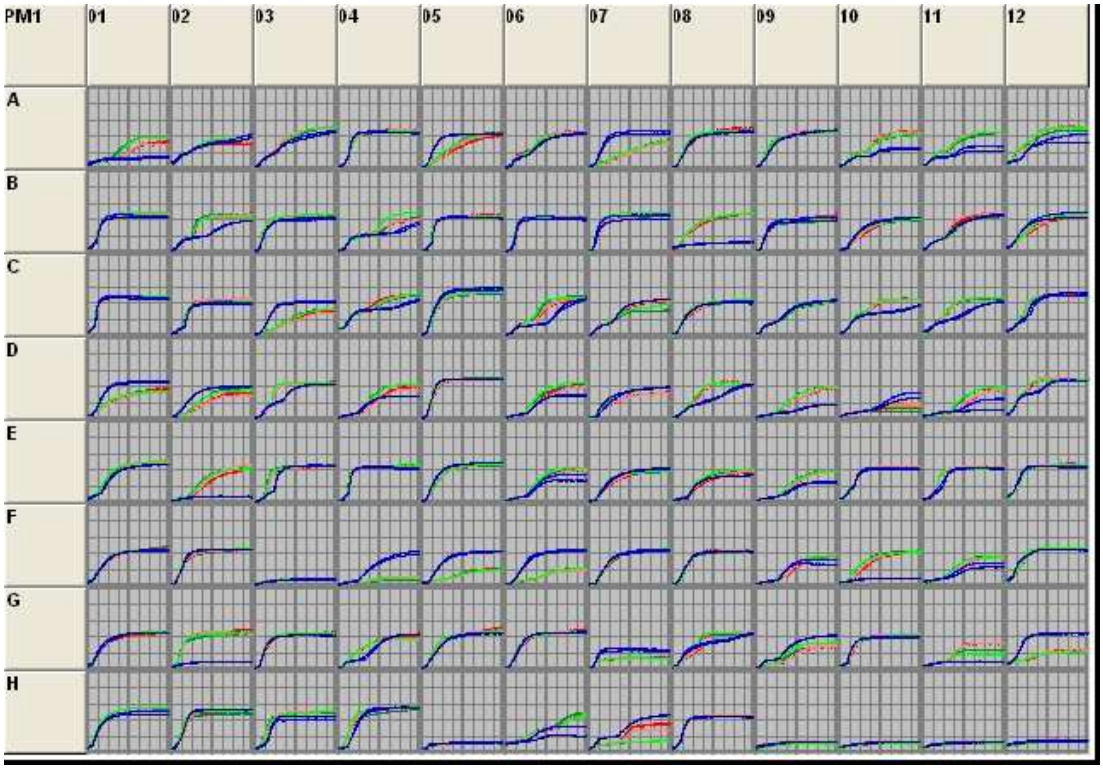
Appendix 5: Microarray (Biolog) data

The graphs below show the results of the metabolic assay for each Phenotype Microarray (PM) plates used. Each data set contains the template of the PM with the respective graph of the results.

Each well displays the kinetics of the utilisation of a specific metabolite (see template). The blue lines represent the metabolite utilisation by *S.* Typhimurium SL1344, the green lines represent the metabolite's utilisation by *S.* Weltevreden C2346 and the red lines represent the metabolite's utilisation by *S.* Weltevreden 10259. Each experiment was conducted in duplicates therefore each well display 2 line per isolates.

PM1 MicroPlate™ Carbon Sources

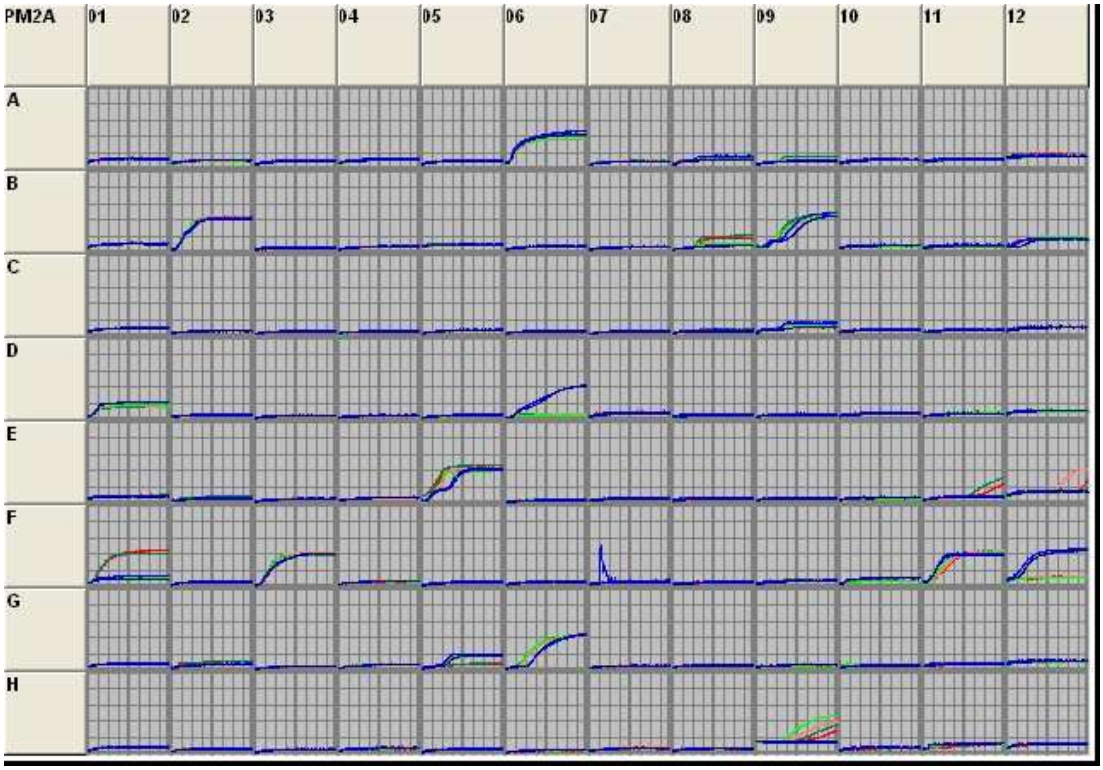
A1 Negative Control	A2 L-Arabinose	A3 N-Acetyl-D-Glucosamine	A4 D-Saccharic Acid	A5 Succinic Acid	A6 D-Galactose	A7 L-Aspartic Acid	A8 L-Proline	A9 D-Alanine	A10 D-Trehalose	A11 D-Mannose	A12 Dulcitol
B1 D-Serine	B2 D-Sorbitol	B3 Glycerol	B4 L-Fucose	B5 D-Glucuronic Acid	B6 D-Gluconic Acid	B7 D,L-α-Glycerol-Phosphate	B8 D-Xylose	B9 L-Lactic Acid	B10 Formic Acid	B11 D-Mannitol	B12 L-Glutamic Acid
C1 D-Glucose-6-Phosphate	C2 D-Galactonic Acid-γ-Lactone	C3 D,L-Malic Acid	C4 D-Ribose	C5 Tween 20	C6 L-Rhamnose	C7 D-Fructose	C8 Acetic Acid	C9 α-D-Glucose	C10 Maltose	C11 D-Melibiose	C12 Thymidine
D-1 L-Asparagine	D2 D-Aspartic Acid	D3 D-Glucosaminic Acid	D4 1,2-Propanediol	D5 Tween 40	D6 α-Keto-Glutaric Acid	D7 α-Keto-Butyric Acid	D8 α-Methyl-D-Galactoside	D9 α-D-Lactose	D10 Lactulose	D11 Sucrose	D12 Uridine
E1 L-Glutamine	E2 m-Tartaric Acid	E3 D-Glucose-1-Phosphate	E4 D-Fructose-6-Phosphate	E5 Tween 80	E6 α-Hydroxy Glutaric Acid-γ-Lactone	E7 α-Hydroxy Butyric Acid	E8 β-Methyl-D-Glucoside	E9 Adonitol	E10 Maltotriose	E11 2-Deoxy Adenosine	E12 Adenosine
F1 Glycyl-L-Aspartic Acid	F2 Citric Acid	F3 m-Inositol	F4 D-Threonine	F5 Fumaric Acid	F6 Bromo Succinic Acid	F7 Propionic Acid	F8 Mucic Acid	F9 Glycolic Acid	F10 Glyoxylic Acid	F11 D-Cellobiose	F12 Inosine
G1 Glycyl-L-Glutamic Acid	G2 Tricarballylic Acid	G3 L-Serine	G4 L-Threonine	G5 L-Alanine	G6 L-Alanyl-Glycine	G7 Acetoacetic Acid	G8 N-Acetyl-β-D-Mannosamine	G9 Mono Methyl Succinate	G10 Methyl Pyruvate	G11 D-Malic Acid	G12 L-Malic Acid
H1 Glycyl-L-Proline	H2 p-Hydroxy Phenyl Acetic Acid	H3 m-Hydroxy Phenyl Acetic Acid	H4 Tyramine	H5 D-Psicose	H6 L-Lyxose	H7 Glucuronamide	H8 Pyruvic Acid	H9 L-Galactonic Acid-γ-Lactone	H10 D-Galacturonic Acid	H11 Phenylethyl-amine	H12 2-Aminoethanol



PM 1 dataset

PM2A MicroPlate™ Carbon Sources

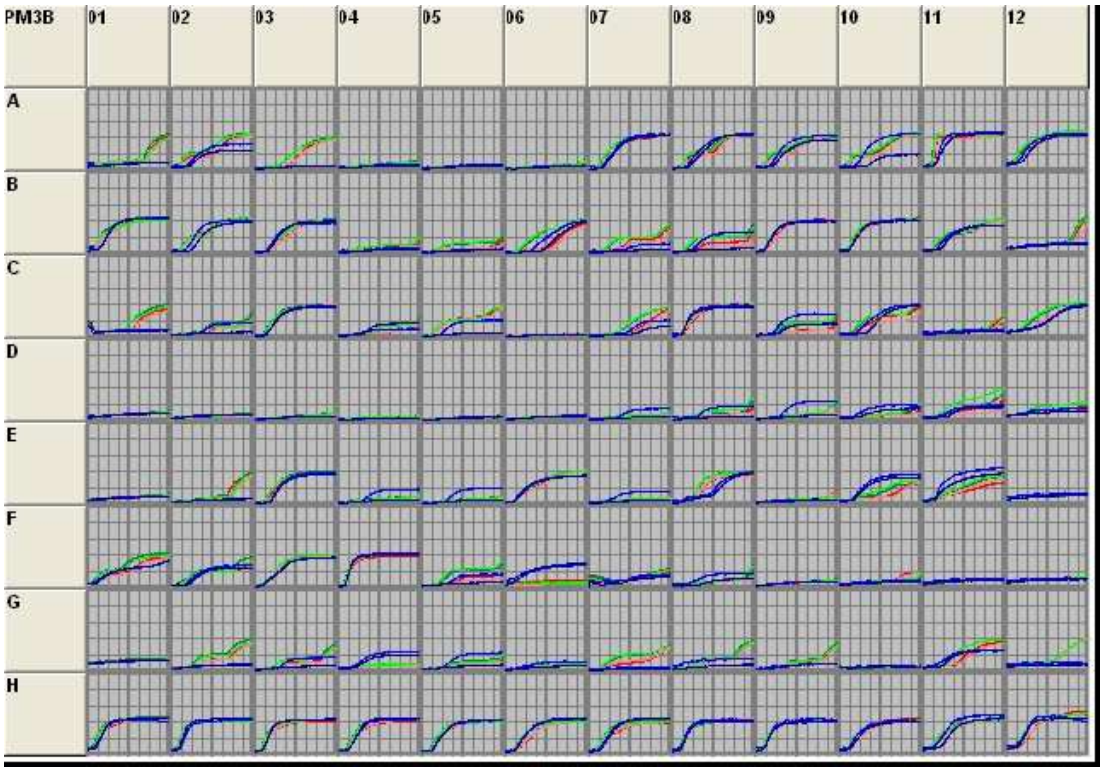
A1 Negative Control	A2 Chondroitin Sulfate C	A3 α-Cyclodextrin	A4 β-Cyclodextrin	A5 γ-Cyclodextrin	A6 Dextrin	A7 Gelatin	A8 Glycogen	A9 Inulin	A10 Laminarin	A11 Mannan	A12 Pectin
B1 N-Acetyl-D-Galactosamine	B2 N-Acetyl-Neuraminic Acid	B3 β-D-Allose	B4 Amygdalin	B5 D-Arabinose	B6 D-Arabitol	B7 L-Arabitol	B8 Arbutin	B9 2-Deoxy-D-Ribose	B10 i-Erythritol	B11 D-Fucose	B12 3-0-β-D-Galacto-pyranosyl-D-Arabinose
C1 Gentiobiose	C2 L-Glucose	C3 Lactitol	C4 D-Melezitose	C5 Maltitol	C6 a-Methyl-D-Glucoside	C7 β-Methyl-D-Galactoside	C8 3-Methyl Glucose	C9 β-Methyl-D-Glucuronic Acid	C10 α-Methyl-D-Mannoside	C11 β-Methyl-D-Xyloside	C12 Palatinose
D1 D-Raffinose	D2 Salicin	D3 Sedoheptulosan	D4 L-Sorbose	D5 Stachyose	D6 D-Tagatose	D7 Turanose	D8 Xylitol	D9 N-Acetyl-D-Glucosaminitol	D10 γ-Amino Butyric Acid	D11 δ-Amino Valeric Acid	D12 Butyric Acid
E1 Capric Acid	E2 Caproic Acid	E3 Citraconic Acid	E4 Citramalic Acid	E5 D-Glucosamine	E6 2-Hydroxy Benzoic Acid	E7 4-Hydroxy Benzoic Acid	E8 β-Hydroxy Butyric Acid	E9 γ-Hydroxy Butyric Acid	E10 a-Keto-Valeric Acid	E11 Itaconic Acid	E12 5-Keto-D-Gluconic Acid
F1 D-Lactic Acid Methyl Ester	F2 Malonic Acid	F3 Melibionic Acid	F4 Oxalic Acid	F5 Oxalomalic Acid	F6 Quinic Acid	F7 D-Ribono-1,4-Lactone	F8 Sebacic Acid	F9 Sorbic Acid	F10 Succinamic Acid	F11 D-Tartaric Acid	F12 L-Tartaric Acid
G1 Acetamide	G2 L-Alaninamide	G3 N-Acetyl-L-Glutamic Acid	G4 L-Arginine	G5 Glycine	G6 L-Histidine	G7 L-Homoserine	G8 Hydroxy-L-Proline	G9 L-Isoleucine	G10 L-Leucine	G11 L-Lysine	G12 L-Methionine
H1 L-Ornithine	H2 L-Phenylalanine	H3 L-Pyroglutamic Acid	H4 L-Valine	H5 D,L-Carnitine	H6 Sec-Butylamine	H7 D,L-Octopamine	H8 Putrescine	H9 Dihydroxy Acetone	H10 2,3-Butanediol	H11 2,3-Butanone	H12 3-Hydroxy 2-Butanone



PM 2 dataset

PM3B MicroPlate™ Nitrogen Sources

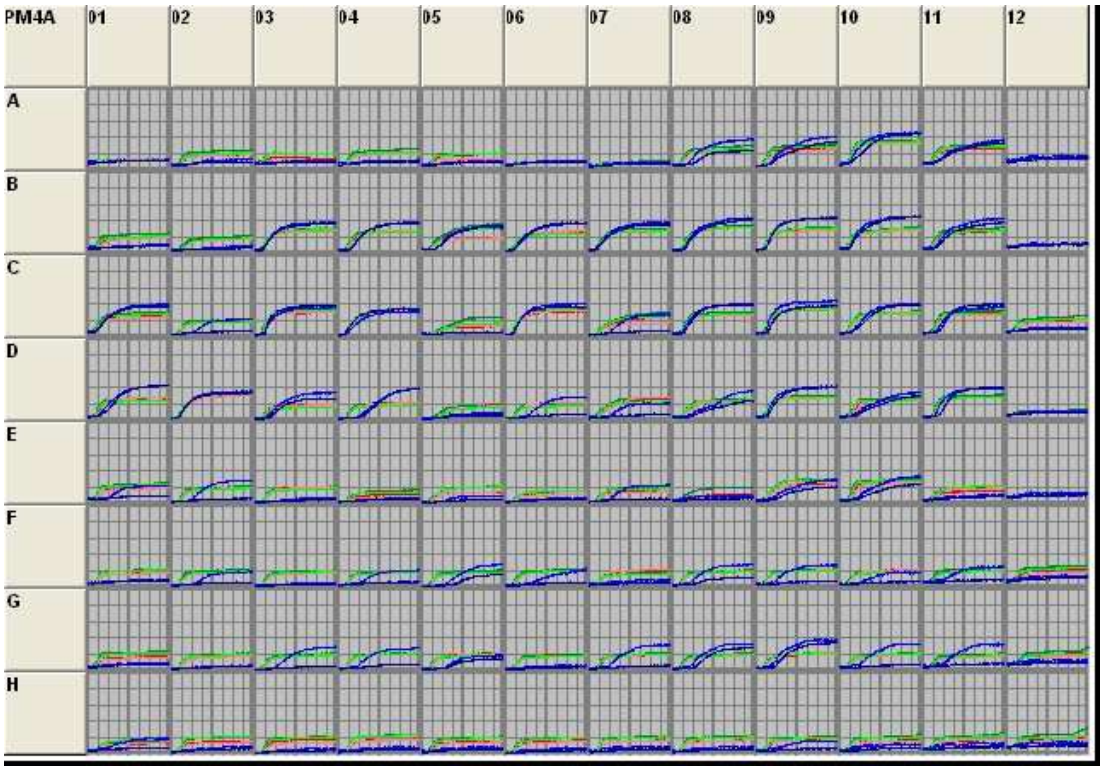
A1 Negative Control	A2 Ammonia	A3 Nitrite	A4 Nitrate	A5 Urea	A6 Biuret	A7 L-Alanine	A8 L-Arginine	A9 L-Asparagine	A10 L-Aspartic Acid	A11 L-Cysteine	A12 L-Glutamic Acid
B1 L-Glutamine	B2 Glycine	B3 L-Histidine	B4 L-Isoleucine	B5 L-Leucine	B6 L-Lysine	B7 L-Methionine	B8 L-Phenylalanine	B9 L-Proline	B10 L-Serine	B11 L-Threonine	B12 L-Tryptophan
C1 L-Tyrosine	C2 L-Valine	C3 D-Alanine	C4 D-Asparagine	C5 D-Aspartic Acid	C6 D-Glutamic Acid	C7 D-Lysine	C8 D-Serine	C9 D-Valine	C10 L-Citrulline	C11 L-Homoserine	C12 L-Ornithine
D-1 N-Acetyl-L-Glutamic Acid	D2 N-Phthaloyl-L-Glutamic Acid	D3 L-Pyroglutamic Acid	D4 Hydroxylamine	D5 Methylamine	D6 N-Amylamine	D7 N-Butylamine	D8 Ethylamine	D9 Ethanolamine	D10 Ethylenediamine	D11 Putrescine	D12 Agmatine
E1 Histamine	E2 β-Phenylethyl-amine	E3 Tyramine	E4 Acetamide	E5 Formamide	E6 Glucuronamide	E7 D,L-Lactamide	E8 D-Glucosamine	E9 D-Galactosamine	E10 D-Mannosamine	E11 N-Acetyl-D-Glucosamine	E12 N-Acetyl-D-Galactosamine
F1 N-Acetyl-D-Mannosamine	F2 Adenine	F3 Adenosine	F4 Cytidine	F5 Cytosine	F6 Guanine	F7 Guanosine	F8 Thymine	F9 Thymidine	F10 Uracil	F11 Uridine	F12 Inosine
G1 Xanthine	G2 Xanthosine	G3 Uric Acid	G4 Alloxan	G5 Allantoin	G6 Parabanic Acid	G7 D,L-α-Amino-N-Butyric Acid	G8 γ-Amino-N-Butyric Acid	G9 ε-Amino-N-Caproic Acid	G10 D,L-α-Amino-Caprylic Acid	G11 δ-Amino-N-Valeric Acid	G12 α-Amino-N-Valeric Acid
H1 Ala-Asp	H2 Ala-Gln	H3 Ala-Glu	H4 Ala-Gly	H5 Ala-His	H6 Ala-Leu	H7 Ala-Thr	H8 Gly-Asn	H9 Gly-Gln	H10 Gly-Glu	H11 Gly-Met	H12 Met-Ala



PM 3 dataset

PM4A MicroPlate™ Phosphorus and Sulfur Sources

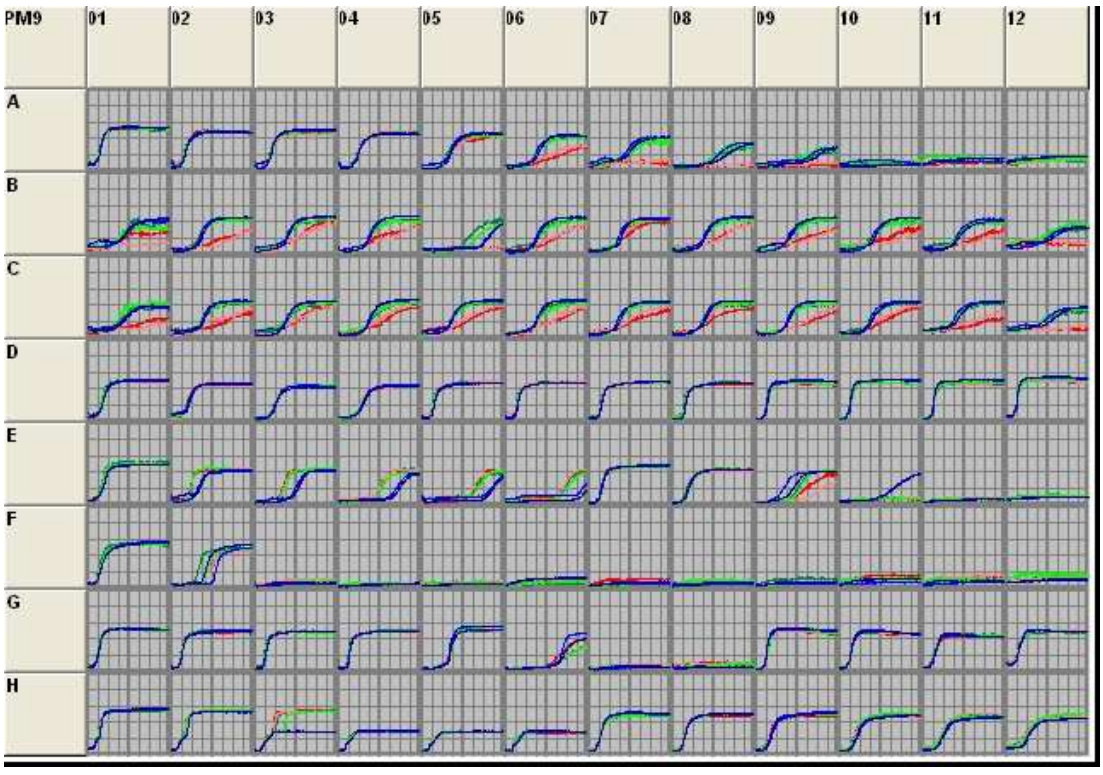
A1 Negative Control	A2 Phosphate	A3 Pyrophosphate	A4 Trimeta-phosphate	A5 Tripoly-phosphate	A6 Triethyl Phosphate	A7 Hypophosphite	A8 Adenosine-2'-monophosphate	A9 Adenosine-3'-monophosphate	A10 Adenosine-5'-monophosphate	A11 Adenosine-2',3'-cyclic monophosphate	A12 Adenosine-3',5'-cyclic monophosphate
B1 Thiophosphate	B2 Dithiophosphate	B3 D,L-α-Glycerol Phosphate	B4 β-Glycerol Phosphate	B5 Carbamyl Phosphate	B6 D-2-Phospho-Glyceric Acid	B7 D-3-Phospho-Glyceric Acid	B8 Guanosine-2'-monophosphate	B9 Guanosine-3'-monophosphate	B10 Guanosine-5'-monophosphate	B11 Guanosine-2',3'-cyclic monophosphate	B12 Guanosine-3',5'-cyclic monophosphate
C1 Phosphoenol Pyruvate	C2 Phospho-Glycolic Acid	C3 D-Glucose-1-Phosphate	C4 D-Glucose-6-Phosphate	C5 2-Deoxy-D-Glucose 6-Phosphate	C6 D-Glucosamine-6-Phosphate	C7 6-Phospho-Gluconic Acid	C8 Cytidine-2'-monophosphate	C9 Cytidine-3'-monophosphate	C10 Cytidine-5'-monophosphate	C11 Cytidine-2',3'-cyclic monophosphate	C12 Cytidine-3',5'-cyclic monophosphate
D1 D-Mannose-1-Phosphate	D2 D-Mannose-6-Phosphate	D3 Cysteamine-S-Phosphate	D4 Phospho-L-Arginine	D5 O-Phospho-D-Serine	D6 O-Phospho-L-Serine	D7 O-Phospho-L-Threonine	D8 Uridine-2'-monophosphate	D9 Uridine-3'-monophosphate	D10 Uridine-5'-monophosphate	D11 Uridine-2',3'-cyclic monophosphate	D12 Uridine-3',5'-cyclic monophosphate
E1 O-Phospho-D-Tyrosine	E2 O-Phospho-L-Tyrosine	E3 Phosphocreatine	E4 Phosphoryl Choline	E5 O-Phosphoryl-Ethanolamine	E6 Phosphono Acetic Acid	E7 2-Aminoethyl Phosphonic Acid	E8 Methylene Diphosphonic Acid	E9 Thymidine-3'-monophosphate	E10 Thymidine-5'-monophosphate	E11 Inositol Hexaphosphate	E12 Thymidine 3',5'-cyclic monophosphate
F1 Negative Control	F2 Sulfate	F3 Thiosulfate	F4 Tetrathionate	F5 Thiophosphate	F6 Dithiophosphate	F7 L-Cysteine	F8 D-Cysteine	F9 L-Cysteinyl-Glycine	F10 L-Cysteic Acid	F11 Cysteamine	F12 L-Cysteine Sulfinic Acid
G1 N-Acetyl-L-Cysteine	G2 S-Methyl-L-Cysteine	G3 Cystathionine	G4 Lanthionine	G5 Glutathione	G6 D,L-Ethionine	G7 L-Methionine	G8 D-Methionine	G9 Glycyl-L-Methionine	G10 N-Acetyl-D,L-Methionine	G11 L-Methionine Sulfoxide	G12 L-Methionine Sulfone
H1 L-Djenkolic Acid	H2 Thiourea	H3 1-Thio-β-D-Glucose	H4 D,L-Lipoamide	H5 Taurocholic Acid	H6 Taurine	H7 Hypotaurine	H8 p-Amino Benzene Sulfonic Acid	H9 Butane Sulfonic Acid	H10 2-Hydroxyethane Sulfonic Acid	H11 Methane Sulfonic Acid	H12 Tetramethylene Sulfone



PM 4 dataset

PM9 MicroPlate™ Osmolytes

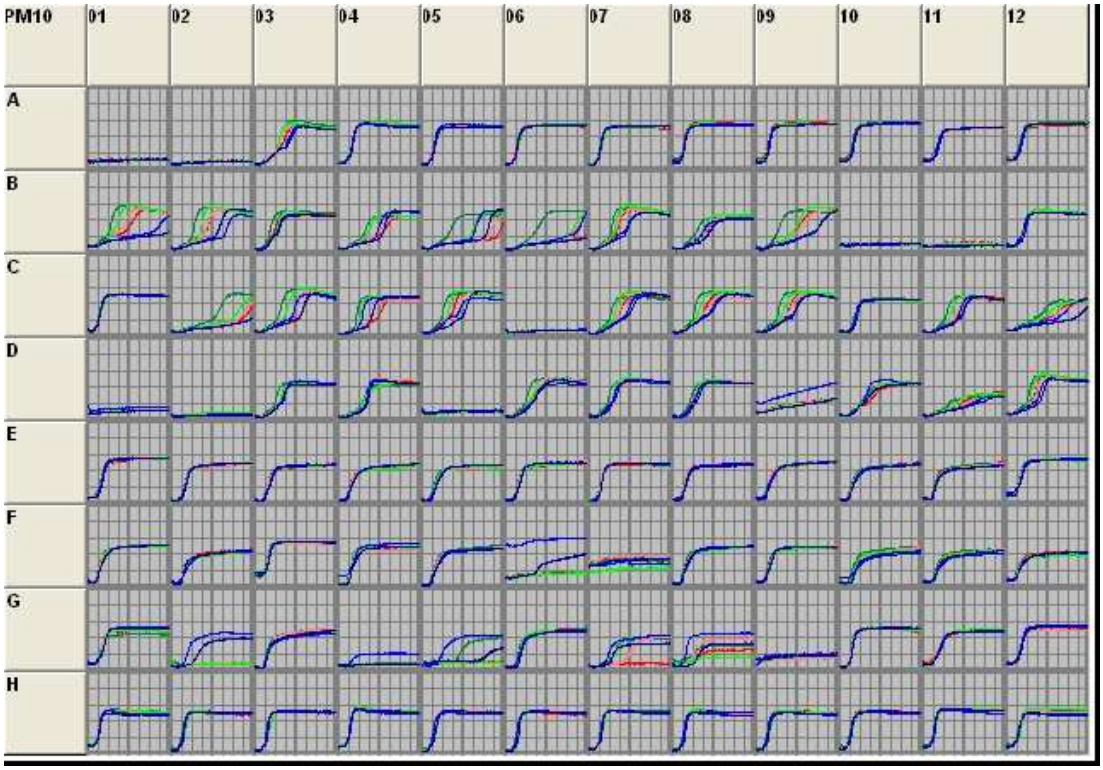
A1 NaCl 1%	A2 NaCl 2%	A3 NaCl 3%	A4 NaCl 4%	A5 NaCl 5%	A6 NaCl 5.5%	A7 NaCl 6%	A8 NaCl 6.5%	A9 NaCl 7%	A10 NaCl 8%	A11 NaCl 9%	A12 NaCl 10%
B1 NaCl 6%	B2 NaCl 6% + Betaine	B3 NaCl 6% + N-N Dimethyl Glycine	B4 NaCl 6% + Sarcosine	B5 NaCl 6% + Dimethyl sulphonyl propionate	B6 NaCl 6% + MOPS	B7 NaCl 6% + Ectoine	B8 NaCl 6% + Choline	B9 NaCl 6% + Phosphoryl Choline	B10 NaCl 6% + Creatine	B11 NaCl 6% + Creatinine	B12 NaCl 6% + L- Carnitine
C1 NaCl 6% + KCl	C2 NaCl 6% + L-Proline	C3 NaCl 6% + N-Acetyl L-Glutamine	C4 NaCl 6% + β-Glutamic Acid	C5 NaCl 6% + γ–Amino -N- Butyric Acid	C6 NaCl 6% + Glutathione	C7 NaCl 6% + Glycerol	C8 NaCl 6% + Trehalose	C9 NaCl 6% + Trimethylamine- N-oxide	C10 NaCl 6% + Trimethylamine	C11 NaCl 6% + Octopine	C12 NaCl 6% + Trigonelline
D1 Potassium chloride 3%	D2 Potassium chloride 4%	D3 Potassium chloride 5%	D4 Potassium chloride 6%	D5 Sodium sulfate 2%	D6 Sodium sulfate 3%	D7 Sodium sulfate 4%	D8 Sodium sulfate 5%	D9 Ethylene glycol 5%	D10 Ethylene glycol 10%	D11 Ethylene glycol 15%	D12 Ethylene glycol 20%
E1 Sodium formate 1%	E2 Sodium formate 2%	E3 Sodium formate 3%	E4 Sodium formate 4%	E5 Sodium formate 5%	E6 Sodium formate 6%	E7 Urea 2%	E8 Urea 3%	E9 Urea 4%	E10 Urea 5%	E11 Urea 6%	E12 Urea 7%
F1 Sodium Lactate 1%	F2 Sodium Lactate 2%	F3 Sodium Lactate 3%	F4 Sodium Lactate 4%	F5 Sodium Lactate 5%	F6 Sodium Lactate 6%	F7 Sodium Lactate 7%	F8 Sodium Lactate 8%	F9 Sodium Lactate 9%	F10 Sodium Lactate 10%	F11 Sodium Lactate 11%	F12 Sodium Lactate 12%
G1 Sodium Phosphate pH 7 20mM	G2 Sodium Phosphate pH 7 50mM	G3 Sodium Phosphate pH 7 100mM	G4 Sodium Phosphate pH 7 200mM	G5 Sodium Benzoate pH 5.2 20mM	G6 Sodium Benzoate pH 5.2 50mM	G7 Sodium Benzoate pH5.2 100mM	G8 Sodium Benzoate pH 5.2 200mM	G9 Ammonium sulfate pH8 10mM	G10 Ammonium sulfate pH 8 20mM	G11 Ammonium sulfate pH 8 50mM	G12 Ammonium sulfate pH8 100mM
H1 Sodium Nitrate 10mM	H2 Sodium Nitrate 20mM	H3 Sodium Nitrate 40mM	H4 Sodium Nitrate 60mM	H5 Sodium Nitrate 80mM	H6 Sodium Nitrate 100mM	H7 Sodium Nitrite 10mM	H8 Sodium Nitrite 20mM	H9 Sodium Nitrite 40mM	H10 Sodium Nitrite 60mM	H11 Sodium Nitrite 80mM	H12 Sodium Nitrite 100mM



PM 9 dataset

PM10 MicroPlate™ pH

A1 pH 3.5	A2 pH 4	A3 pH 4.5	A4 pH 5	A5 pH 5.5	A6 pH 6	A7 pH 7	A8 pH 8	A9 pH 8.5	A10 pH 9	A11 pH 9.5	A12 pH 10
B1 pH 4.5	B2 pH 4.5 + L-Alanine	B3 pH 4.5 + L-Arginine	B4 pH 4.5 + L-Asparagine	B5 pH 4.5 + L-Aspartic Acid	B6 pH 4.5 + L-Glutamic Acid	B7 pH 4.5 + L-Glutamine	B8 pH 4.5 + Glycine	B9 pH 4.5 + L-Histidine	B10 pH 4.5 + L-Isoleucine	B11 pH 4.5 + L-Leucine	B12 pH 4.5 + L-Lysine
C1 pH 4.5 + L-Methionine	C2 pH 4.5 + L-Phenylalanine	C3 pH 4.5 + L-Proline	C4 pH 4.5 + L-Serine	C5 pH 4.5 + L-Threonine	C6 pH 4.5 + L-Tryptophan	C7 pH 4.5 + L-Citrulline	C8 pH 4.5 + L-Valine	C9 pH 4.5 + Hydroxy-L-Proline	C10 pH 4.5 + L-Ornithine	C11 pH 4.5 + L-Homoarginine	C12 pH 4.5 + L-Homoserine
D-1 pH 4.5 + Anthranilic Acid	D2 pH 4.5 + L-Norleucine	D3 pH 4.5 + L-Norvaline	D4 pH 4.5 + α- Amino-N-Butyric Acid	D5 pH 4.5 + p-Amino-Benzoic Acid	D6 pH 4.5 + L-Cysteic Acid	D7 pH 4.5 + D-Lysine	D8 pH 4.5 + 5-Hydroxy Lysine	D9 pH 4.5 + 5-Hydroxy Tryptophan	D10 pH 4.5 + D,L-Diamino pimelic Acid	D11 pH 4.5 + Trimethylamine-N-oxide	D12 pH 4.5 + Urea
E1 pH 9.5	E2 pH 9.5 + L-Alanine	E3 pH 9.5 + L-Arginine	E4 pH 9.5 + L-Asparagine	E5 pH 9.5 + L-Aspartic Acid	E6 pH 9.5 + L-Glutamic Acid	E7 pH 9.5 + L-Glutamine	E8 pH 9.5 + Glycine	E9 pH 9.5 + L-Histidine	E10 pH 9.5 + L-Isoleucine	E11 pH 9.5 + L-Leucine	E12 pH 9.5 + L-Lysine
F1 pH 9.5 + L-Methionine	F2 pH 9.5 + L-Phenylalanine	F3 pH 9.5 + L-Proline	F4 pH 9.5 + L-Serine	F5 pH 9.5 + L-Threonine	F6 pH 9.5 + L-Tryptophan	F7 pH 9.5 + L-Tyrosine	F8 pH 9.5 + L-Valine	F9 pH 9.5 + Hydroxy-L-Proline	F10 pH 9.5 + L-Ornithine	F11 pH 9.5 + L-Homoarginine	F12 pH 9.5 + L-Homoserine
G1 pH 9.5 + Anthranilic acid	G2 pH 9.5 + L-Norleucine	G3 pH 9.5 + L-Norvaline	G4 pH 9.5 + Agmatine	G5 pH 9.5 + Cadaverine	G6 pH 9.5 + Putrescine	G7 pH 9.5 + Histamine	G8 pH 9.5 + Phenylethylamine	G9 pH 9.5 + Tyramine	G10 pH 9.5 + Creatine	G11 pH 9.5 + Trimethylamine-N-oxide	G12 pH 9.5 + Urea
H1 X-Caprylate	H2 X-α-D-Glucoside	H3 X-β-D-Glucoside	H4 X-α-D-Galactoside	H5 X-β-D-Galactoside	H6 X-α- D-Glucuronide	H7 X-β- D-Glucuronide	H8 X-β-D-Glucosaminide	H9 X-β-D-Galactosaminide	H10 X-α-D-Mannoside	H11 X-PO4	H12 X-SO4



PM 10 dataset